

AMERICAN HISTORICAL ASSOCIATION

INVESTIGATION OF THE
SOCIAL STUDIES IN THE SCHOOLS



STAFF

A. C. KREY

Chairman, Director of the Investigation

G. S. COUNTS

Research Director

W. G. KIMMEL

Executive Secretary

T. L. KELLEY

Psychologist, Advisor on Tests

COMMISSION ON DIRECTION

FRANK W. BALLOT, Superintendent of Schools, Washington, D. C.

CHARLES A. BEARD, formerly Professor of Politics, Columbia University; author of many books in the fields of history and politics.

ISAIAH BOWMAN, Director, American Geographical Society of New York; President of the International Geographical Union.

ADA COMSTOCK, President of Radcliffe College.

GEORGE S. COUNTS, Professor of Education, Teachers College, Columbia University.

AVERY O. CRAVEN, Professor of History, University of Chicago.

EDMUND E. DAY, formerly Dean of School of Business Administration, University of Michigan; now Director of Social Sciences, Rockefeller Foundation.

GUY STANTON FORD, Professor of History, Dean of the Graduate School, University of Minnesota.

CARLTON J. H. HAYES, Professor of History, Columbia University.

ERNEST HORN, Professor of Education, University of Iowa.

HENRY JOHNSON, Professor of History, Teachers College, Columbia University.

A. C. KREY, Professor of History, University of Minnesota.

LEON C. MARSHALL, Institute for the Study of Law, Johns Hopkins University.

CHARLES E. MERRIAM, Professor of Political Science, University of Chicago.

JESSE H. NEWLON, Professor of Education, Teachers College, Columbia University; Director of Lincoln Experimental School.

JESSE F. STEINER, Professor of Sociology, University of Washington.

TESTS AND MEASUREMENTS
IN THE
SOCIAL SCIENCES

REPORT OF THE COMMISSION
ON THE SOCIAL STUDIES
PART IV

TESTS AND MEASUREMENTS
IN THE
SOCIAL SCIENCES

BY

TRUMAN L. KELLEY

Graduate School of Education, Harvard University

AND

A. C. KREY

Professor of History, University of Minnesota

CHARLES SCRIBNER'S SONS

NEW YORK

CHICAGO

BOSTON

ATLANTA

SAN FRANCISCO

DALLAS

4467

COPYRIGHT, 1934, BY
CHARLES SCRIBNER'S SONS

Printed in the United States of America

*All rights reserved. No part of this book
may be reproduced in any form without
the permission of Charles Scribner's Sons*

A



PREFACE

The present work is the fourth in the series of volumes which constitute the report of the Commission on the Social Studies in the Schools. The first three volumes present successively a view of the problem, its historical background, and the concern of the public in the problem. This, the fourth, presents the experience of the Commission in its study of objective tests and measurements in the field of the social studies.

The work of test construction was at first directly under the charge of the Commission as a whole. There was at that time some hope that the whole field of social science instruction might be submitted to objective measurement with important implications for each division of the work. There was also some awareness of the long time required for the preliminary analytical work. For these reasons the Commission provided early for this work on tests. Early reports of progress, however, indicated that the work of test construction involved a high degree of technical knowledge. The Commission accordingly appointed an Advisory Committee consisting of Frank W. Ballou, Chairman, Isaiah Bowman, Howard C. Hill, Ernest Horn, Henry Johnson, A. C. Krey, and Ben D. Wood to have immediate direction of test construction. Truman L. Kelley, psychologist, advised the Committee. This Committee considered the tests constructed by Edgar B. Wesley, Mary G. Kelty, Marion Clark, and those of Edith P. Parker and R D Calkins. Its members were also consulted for advice on the portions of this volume contributed by T. L. Kelley and A. C. Krey. Obviously, therefore, the Committee has only a partial responsibility for the material here presented. The authors of specific sections as given in the Table of Contents are responsible for such sections and the Chairman of the Commission is responsible for any matter not otherwise credited.

It will be difficult, if not impossible, to acknowledge all the help received in the preparation of this volume. Numerous tests were analyzed; some already standardized, others in earlier stages of development, some published, others in less permanent form, each the product of more or less cooperation. The analytical work of find-

PREFACE

ing items for use in tests, the independent checking of these items, the judgment of the items by consensus of opinion, the preliminary construction of test questions, critical judgment of such questions, the trial of these questions under a sufficiently varied set of school conditions, as well as the evaluation of statistical findings, all required the assistance of a number of scholars and teachers. This assistance has been rendered at intervals over a period of nearly five years. Under the circumstances, the chance that some even of those who aided directly have been overlooked is too great to justify the hope that all assistance has been acknowledged. We can only hope that such omissions will be discovered before the final volume of the Commission's report is published so that due acknowledgment may be made at that time.

The late Frances E. Baldwin, O. M. Dickerson, Mary Gold, Donald L. McMurry, Joseph R. Strayer, Edith E. Ware, and Edgar B. Wesley were attached to the staff of the Investigation for varying periods and devoted much of their attention to this work. Dorothy Bovee, Gladys Boyington, Dorothy Houston, Herman Muelder, Renata Pecinovsky, Lester N. Recktenwald, William Shepherd, Garland Taylor, and Olga Wold, teachers of varied experience who were continuing their graduate study, have aided in this work for definite periods of time.

We are indebted for various forms of voluntary assistance to the following of our colleagues throughout the country: H. S. Anderson, John Aseltine, Elizabeth Baker, R. Bell, Robert C. Binkley, H. H. Bixler, Ernest L. Bogart, Witt Bowden, D. S. Brainard, Harold H. Burbank, Kathryn E. C. Carrigan, J. H. Carman, Thomas N. Carver, P. Eugene Davenport, Edgar Dawson, Virgil E. Dickson, Evalyn Dixon, Theodore Estabrook, F. R. Fairchild, Halley Farmer, Frederic B. Garver, John M. Gaus, Alice N. Gibbons, Charles E. Greene, J. John Halverson, P. M. Hamer, J. R. Harper, W. H. Hathaway, Mark M. Heald, E. S. Henry, J. D. Hicks, Grace Hotchkiss, R. O. Hughes, Horace Kidger, Von E. Kniseley, Jessie La Salle, Benjamin E. Lippincott, John Lund, Paul E. Lutz, I. N. Madsen, Carl A. Melby, Floyd B. Moe, DeWitt S. Morgan, Eugene A. Nifenecker, A. H.

PREFACE

Noyes, C. W. Odell, Lawrence Packard, Clarence Perkins, J. E. Pomfret, Arthur Repp, Edgar E. Robinson, Theodore Saam, Allan Saunders, Caroline Shanks, J. L. Sharfman, Richard H. Shryock, Russell A. Stevenson, Edna H. Stone, E. J. Sweeney, Lida Lee Tall, V. T. Thayer, Brian E. Tomlinson, Charles F. Towne, Alice Felt Tyler, Laura F. Ullrick, R. S. Vaile, E. M. Violette, Carleton Washburne, W. P. Webb, Raymond Werner, Benjamin W. Wheeler, A. C. Wilgus, Malcolm Willey, Fremont P. Wirth, Dean A. Worcester, and John Wuorinen.

Many school systems have aided in the trial stages of the several tests. As the tests were of an experimental nature and bore no direct relationship to courses as conducted in any of the schools, they could not be used for general examination purposes. The trial of the tests, therefore, was confined to limited groups of pupils in any one school system and the whole picture was completed by piecing together the returns from scattered groups. For their assistance under these circumstances the Commission feels especially grateful to the school authorities of Americus, Kansas; Anoka, Minnesota; Asheville, North Carolina; Atlanta, Georgia; Atlantic City, New Jersey; Avondale, North Carolina; Benson, North Carolina; Beverley, Massachusetts; Biltmore, North Carolina; Bonlee, North Carolina; Cambridge, Massachusetts; Caroleen, North Carolina; Chadbourne, North Carolina; Cliffside, North Carolina; Clinton, North Carolina; Columbus, Ohio; Charleston, Massachusetts; Charleston, West Virginia; Hyde Park High School, Chicago, Illinois; Delgado, North Carolina; Denver, Colorado; Dearborn, Michigan; Elgin, Illinois; Emporia, Kansas; Excelsior, Minnesota; Falcon, North Carolina; Faribault, Minnesota; Glassboro, New Jersey; Goldston, North Carolina; Grosse Point, Michigan; Haddon Heights, New Jersey; Henderson, North Carolina; Hendersonville, North Carolina; Highland Park, Michigan; Horace Mann School, New York City; Houston, Texas; Indianapolis, Indiana; Kansas City, Missouri; Kenosha, Wisconsin; Lattimore, North Carolina; Lawrence School, Hewlett, New York; Long Beach, California; Lyons, New York; Maiden, North Carolina; Milwaukee, Wisconsin;

PREFACE

Minneapolis, Minnesota; Monroe, North Carolina; New Orleans, Louisiana; Newton, North Carolina; Newtonville, Massachusetts; New York City; Norwalk, Connecticut; Oakland, California; Old Ford, North Carolina; Overbrook, Pennsylvania; Owatonna, Minnesota; Pantego, North Carolina; Pen Yan, New York; Philadelphia, Pennsylvania; Pittsboro, North Carolina; Pittsburgh, Pennsylvania; Providence, Rhode Island; Reading, Massachusetts; Roanoke, Virginia; Rochester, New York; Rock Springs, Georgia; Roseboro, North Carolina; Rutherfordton, North Carolina; Sacramento, California; St. Louis, Missouri; St. Paul, Minnesota; San Diego, California; Sanford, North Carolina; Seattle, Washington; Shakopee, Minnesota; South Hamilton, Massachusetts; Summit, New Jersey; Turkey, North Carolina; Urbana, Illinois; Washington, D. C.; Washington, North Carolina; Wenham, Massachusetts; Whiteville, North Carolina; Wilmette, Illinois; Wilmington, North Carolina; Winnetka, Illinois; Wrightsville, Georgia.

We are also indebted for assistance on one or more of the tests to the following colleges: Duke University, Durham, North Carolina; State Teachers College, Emporia, Kansas; University of California, Berkeley; State Teachers College, Flagstaff, Arizona; Teachers College, Columbia University, New York City; Harvard University, Cambridge, Massachusetts; Hunter College, New York City; University of Louisiana, Baton Rouge; State College for Women, Montevallo, Alabama; State Normal School, Lewiston, Montana; University of Illinois, Urbana; University of Minnesota, Minneapolis; University of Nebraska, Lincoln; New York University, New York City; University of North Dakota, Grand Forks; Ohio State University, Columbus; Princeton University, Princeton, New Jersey; St. Olaf College, Northfield, Minnesota; State Teachers College, St. Cloud, Minnesota; Stanford University, Palo Alto, California; University of Tennessee, Knoxville; State Normal School, Towson, Maryland; State Teachers College, Virginia, Minnesota; George Peabody College for Teachers, Nashville, Tennessee.

Grateful acknowledgment for criticism and help on the technical aspects of the problem is also made to the following: H. R. Ander-

PREFACE

son, Philip A. Boyer, Prudence Cutright, Cecile White Flemming, Mary Harden, William C. Knoelk, E. F. Lindquist, Caroline Pfaff, Irene M. Steele, O. M. Stephenson, Lena C. Van Bibber, Howard E. Wilson, and Clifford Woody.

It would be ingratitude indeed to leave the impression even by implication that the many individuals and institutions which so kindly gave assistance were in any way responsible for any shortcomings which may appear in this work. Their help was given at various stages and on various parts of the work herein described. Few of them had any share in planning the work and fewer still in passing upon the conclusions reached. It is even possible that some of them would find themselves at variance with a number of the conclusions herein set forth. For these reasons it seems desirable to absolve them from any such responsibility, while, at the same time, acknowledging with gratitude the assistance which they have rendered.

Each of the contributions in this volume bears the name of the individual immediately responsible for its statements. Any evaluation of these contributions must, of necessity, be made in the light of the background from which the individual approached his task. Several of the contributions are made by scholars in educational measurement and psychology who have had very little occasion to proceed far, or to remain in close touch, with advanced work in social science subjects. Other contributions are made by teachers at various levels of school instruction—the elementary school, the high school, and college. These teachers have had various degrees of experience with the technical aspects of the new type test. Statistical work and interpretation on all of these tests was done under the immediate direction of T. L. Kelley. While an effort was made, through the assistance of T. L. Kelley and A. C. Krey, to bring to bear on each of these problems the help of advanced work both in technique and subject matter, each individual contributor was free to make such use of this aid as seemed to him wise and pertinent. As some of the contributions show, there were differences of opinion as to the value of the advice received; and in some instances there is evidence

PREFACE

of misunderstanding. The conclusions presented in the separate articles are therefore those of the individual contributors and must be so regarded. This is equally true of the statement of general conclusions from the point of view of instruction in the social science subjects. The Commission has indicated separately its own judgment on the value of the new type test in its volume of Conclusions.

The writer of this preface would like to take advantage of this opportunity to express his own gratitude for the many helpful suggestions and kindly criticisms which he received on his portion of this volume from H. R. Anderson, Frank W. Ballou, Isaiah Bowman, Edmund E. Day, M. E. Haggerty, Edna Heidebreder, Howard C. Hill, Ernest Horn, Henry Johnson, John B. Johnston, E. F. Lindquist, Donald G. Paterson, John G. Rockwell, Joseph R. Strayer, David F. Swenson, Edgar B. Wesley, Ben D. Wood, Clifford Woody, and C. S. Yoakum. Ernest Horn, H. R. Anderson, and E. F. Lindquist aided the writer on his preliminary as well as later draft of the material. Above all, the writer is indebted to Truman L. Kelley, not only for reading all the versions, but for his unfailing patience in instructing him in the more technical aspects of the problem during five years of close association.

A. C. KREY.

CONTENTS

CHAPTER	PAGE
I. INTRODUCTION—A. C. KREY	1-119
PROCEDURE ADOPTED	1-4
I. THE EFFORT TO TEST UNDERSTANDING	5-52
II. HISTORICAL-MINDEDNESS AND CULTURE PATTERNS	53-56
III. SKILLS	57-74
IV. INTERESTS AND ATTITUDES	75-94
V. COMPARISON OF OLD AND NEW TYPE TESTS	95-101
V. CONCLUSIONS	102-112
VI. PROBABLE LEARNING PROCESS IN SOCIAL SCIENCE	113-119
II. PLAN AND METHOD OF TEST CONSTRUCTION	120-153
ORGANIZATION OF TEST ACTIVITIES—A. C. KREY	120-128
GENERAL MEASUREMENT PLAN—T. L. KELLEY	129-152
III. TESTING OF UNDERSTANDING	153-233
PLAN ADOPTED	153-154
A STUDY IN THE LEARNING OF THE FUNDAMENTAL SPECIAL VOCABULARY OF HISTORY—LUELLE COLE PRESSEY	155-218
WESLEY TEST IN SOCIAL TERMS—EDGAR B. WESLEY	219-226
KELTY-MOORE TEST OF CONCEPTS IN THE SOCIAL STUDIES—MARY G. KELTY AND NELLE E. MOORE	227-233

CONTENTS

CHAPTER	PAGE
IV. TESTING OF SKILLS	234-339
OBJECTIVE	234
REPORT OF THE WORK OF THE COMMITTEE ON GEOGRAPHY TESTS—EDITH PUTNAM PARKER	235-296
SUMMARY OF REPORT UPON THE ITEM STUDY OF THE CALKINS-PARKER GEOGRAPHY TESTS—T. L. KELLEY	296-301
THE CONSTRUCTION OF EXERCISES IN THE USE OF HISTORICAL EVIDENCE—MARION CLARK	302-339
V. INTERESTS AND ATTITUDES	340-471
STATEMENT OF PLANNING COMMITTEE	340-341
THE MEASUREMENT OF CERTAIN TRAITS OF CHARACTER—T. L. KELLEY AND M. R. TRABUE	342-437
PARSIMONY IN THE JUDGMENT AND MEASUREMENT OF CHARACTER TRAITS—T. L. KELLEY	403-436
CHEATING IN THE CLASSROOM—A. M. JORDAN	437-471
VI. SUMMARY AND CONCLUSIONS—A. C. KREY	472-485
STATEMENT OF OBJECTIVES	472-474
GRADE PLACEMENT	474-476
NATURAL ORDER OF LEARNING IN THE SOCIAL STUDIES	477-480
NEW-TYPE TEST IN THE SOCIAL SCIENCES	480-485
APPENDIX I	
A DIVERGENT OPINION AS TO THE FUNCTION OF TESTS AND TESTING—T. L. KELLEY	487-501
APPENDIX II	
TERMS IN THE SOCIAL SCIENCES	502-609
BIBLIOGRAPHY	610-633

TESTS AND MEASUREMENTS
IN THE
SOCIAL SCIENCES

CHAPTER I

INTRODUCTION

A. C. KREY

The members of the Commission have felt that the record of our work in the preparation and use of tests should be accompanied by some appraisal of such testing procedure, both as an instrument of investigation and as a possible aid to instruction. The Executive Committee has assigned this task to the Chairman.

Such an appraisal is a highly difficult problem which can scarcely be solved to the complete satisfaction of any group or individual. It involves the judgment of instruments devised by psychologists and other workers in educational measurement for the investigation of learning and instruction in social science subjects. Experts in one field of learning are forced to consider the work of experts in another field. The learning process, one of the chief preoccupations of psychology, must be considered as applied to social science, the special preoccupation of another group of scholars. A philosopher might write an interesting essay on such a problem, but it would probably not be accepted by psychologists and would have little direct value to the teacher in the social sciences. Whether a social scientist may undertake the task with any greater assurance is doubtful.

Psychology is not as yet a definite and unified science yielding incontrovertible results of practical utility to other fields of learning. The problems with which it deals are many. It approaches these problems from a number of different premises. There are several different schools or systems of psychology at present, and individual psychologists draw freely from two or even more of these systems.¹ These schools and individual workers are marked not alone by the particular problems of psychology which they attack

¹Heidbreder, E., *Seven Psychologies*, New York, 1933; Woodworth, R. S., *Contemporary Schools of Psychology*, New York, 1931; Higginson, G. D., *Fields of Psychology*, New York, 1931.

TESTS AND MEASUREMENTS

and the different premises from which they start but nearly as much by the different vocabularies which they use in describing their problems and procedures. All of these psychologists are concerned with some problems of interest to the social scientists, and some problems of social science are of interest to all schools of psychology. To the layman it might appear, therefore, that both psychologist and social scientist should be able to offer a definite contribution to the common task. No one, however, would at present feel qualified to define that contribution in such a way as to satisfy the psychologist and yet be of service to the teacher in social science.

The writer has chosen, therefore, to present his statement more or less as a narrative of the work of test construction during a period of four years, together with such judgments as have occurred to him in watching that development. The trials and errors as well as the achievements have been recorded. Conclusions reached at one stage are shown undergoing revisions at the next. This procedure has been followed in the hope that the reader, whether interested in educational measurement or in teaching the social sciences, may thus share the experience of the test committee and apply to the series of experiences the technical terms which he prefers. This effort will have served its purpose if it calls to the attention of workers in both fields the nature of the problems involved and the possibility and desirability of greater mutual service.

This narrative and appraisal of our work with tests may serve as an introduction to the subject for teachers who, like the writer, are relatively inexperienced in the newer techniques of measurement. Those who are familiar with these techniques may find it more profitable to begin with the more technical material which follows. The determination and details of the plan devised for this phase of the investigation are described later.¹ For the purposes of this introduction we may begin with the scrutiny of tests available in 1929, when this work was started.

¹See Chap. II.

SCRUTINY OF EXISTING TESTS

SCRUTINY OF EXISTING TESTS

The application of the new test technique to the social studies was well under way before this Investigation was begun. Efforts had been directed chiefly to the measurement of results in the elementary grades. Some work had been done at the secondary school levels and a few special projects had been launched in the college field. Enterprising publishers of textbooks, recognizing in the new-type test a coming vogue of the educational world, had begun to develop tests to accompany their textbooks.

Most of the tests available might be classed as tests of specific information or of definitions. Some of the tests combined the two. A few attempted to test "relationship," "thought," and "judgment." A variety of new-type questions were applied in these tests, the true-false, multiple-choice, matching and completion types being most common. Paragraph analysis and some critical exercises were included under the general form of reading-comprehension tests.

A few of the tests were standardized; that is, they had been tried out in an experimental procedure with a sufficient number of pupils to afford a basis for comparison useful for the country as a whole. Such tests were usually published in several equivalent forms, thus affording an opportunity to test progress in a given class.¹

The pedagogical literature expounding and criticizing these tests indicated that the emphasis in their construction had been placed upon the measurement technique.² Both the reliability of the equivalent forms and the validity of the tests as measures of progress rested upon the statistical showing of the results. There was almost no critical discussion of the tests as legitimate measures of knowledge as determined by those who had advanced far in the knowledge of specific subjects.

Several months were spent in the examination of these tests by Dr. Kelley with the assistance of the members of the staff of the

¹See Bibliography. ²This observation has recently been confirmed in a more technical treatment of the matter by E. F. Lindquist and H. R. Anderson: *Achievement Tests in the Social Studies*, *Educational Record*, April, 1933.

TESTS AND MEASUREMENTS

Investigation. It was agreed that the greatest merit of the existing tests lay in their measurement of range of specific information and in their recall of specific verbal definitions. For the purposes of this Investigation, however, this merit was of very limited value. Pupils who had had a course in ancient history might be expected to know a considerable amount of specific information in that field, and students who had had a course in sociology might be expected to know the precise definitions offered in that course. What, however, would be the value of applying such tests to pupils who had had neither ancient history nor sociology? Some, perhaps, but not enough to justify the construction of tests for every course offered in every school system nor the trouble involved in attempting to administer a nation-wide application of such tests. The tests of "relationship," "thought," and "judgment" were found either to depend too closely upon specific subject matter of separate courses or to be too questionable in their assumptions to serve the needs of the Investigation.

It became increasingly clear that the whole range of information and values offered by the social studies program of the country could not be tested within the resources of the Investigation. The Commission must content itself with achieving as much as possible within its limited means. This necessity made it further desirable to focus the efforts of the testing program upon those values which the several subject-matter fields had in common rather than upon their separate contributions. The tentative list¹ of these more nearly common values as drawn up by the planning committee became increasingly important. Existing tests afforded practically no help in this endeavor. The relatively slight progress which the tests of "relationship," "thought," and "judgment" had made toward the measurement of the more intangible values contained in the tentative list of objectives was not very encouraging. Nevertheless it was decided to attempt tests of as many of the types of outcome of instruction contained in the list as possible, provided competent specialists in the social sciences could be found to undertake the actual construction of the tests. As cir-

¹This list appears on pp. 125-127.

THE EFFORT TO TEST UNDERSTANDING

cumstances were to determine, the first of the objectives or outcomes of instruction listed—understanding of important ideas by means of which society functions—received major attention.

I. THE EFFORT TO TEST UNDERSTANDING

Understanding, as the term is used by most teachers of the social sciences, is an intangible value over and above knowledge of specific information or precise verbal definitions. It is one of those values which nearly all teachers recognize, but few have ever undertaken to define. Yet for purposes of test construction it had to be defined, or at least identified, so that the results of tests might afford definite proof of the presence of understanding as defined.

Specific information, referred to as "the facts of history," for example, is distinguished as consisting of those items which are sharply limited in time, place, person, event, and attendant circumstance. A student can commit a great number of such items to memory without attaining understanding. Precise definitions composed of certain words arranged in definite order offer the same difficulty. A student can commit a precise definition to memory without being able to apply it to any given situation. In such cases it becomes a mere verbalism and does not indicate understanding.

It is generally assumed that the student learns in his classes in social science much that will be of use to him as a responsible member of society. Obviously the specific information which he learns in school will all be past facts after he leaves school, whether that information has come to him as a narrative of an event in history or as a group of statistics in other social sciences. The actual life situations which he will face will all be more or less different from those which he has studied in school. The carry-over from school to life, therefore, must consist of something beyond the specific information. It must be something about the information worked over in class that can be applied to other facts. Classification of certain facts, or relationship of facts of certain kinds, it seemed, might furnish the clue to understanding. And if such classifications or relationships were symbolized by terms, these terms could be used in the construction of the new-type test. All such terms, how-

TESTS AND MEASUREMENTS

ever, are capable of some sort of definition and most of them are defined in dictionaries. Ability to repeat the verbal definition of such terms would not prove understanding. Ability to apply the term or its definition correctly would be more convincing, especially if the application were to a new situation not previously met.

The distinction here attempted between information and understanding may offer some difficulties to psychologists who differ among themselves as to the quality and value of verbal symbols on the printed page. Nor does the writer presume to enter into the vexed controversies of the epistemologist about the finer distinctions of idea, sensation, and concept. The distinction grows out of the materials which the social scientist uses and for his purposes is adopted as a practical hypothesis to be tested by results. It might be expressed in another way as the difference between a specific fact and a generalized fact. The statement that Johnny Jones was killed by Mr. Smith's automobile on Brown Street, March 1, 1930, is composed of a series of specific past facts having in themselves no functional value. On the other hand, the statement that crossing a street under similar circumstances is dangerous is a generalized statement of fact which does have functional value. The former may be called an inert fact, the latter a functional fact capable of application to new situations in the future.

The designation "inert" for time- and space-limited facts is an arbitrary one. It is being made to cover a great variety of material which perhaps can not be included under a single label. Recalling Josiah Royce's definition of a community as a group of people who have one fact in common, even the most inert of facts might be used for a purpose and thus to that extent qualify as functional. The only justification for the use of the term in the sense here indicated is to emphasize a distinction between the materials of the social science curriculum, which in themselves have definite limits of time and place, and those materials which are unlimited. This distinction has usually been overlooked in pedagogical discussion. It has been pointed out that the designation of "inert" to items of specific information might actually do much harm by its inherent implication of uselessness, whereas this material may, and, as our

THE EFFORT TO TEST UNDERSTANDING

later experience was to show, actually did, prove to be of vital use to the teacher. Perhaps the suggestion of one critic that "unique" be substituted for "inert," and "general" for "functional," would describe the difference without any damaging connotation. At any rate, the terms should be regarded as purely temporary and subject to important modification in light of later evidence.

It was decided to extract from the matter offered in the social sciences the terms expressing functional facts as the basis for the construction of tests of understanding. To safeguard against the possibility that this distinction was purely an individual or "subjective" idea, five teachers were asked to collect such terms from textbooks used in the social studies from grades four to twelve inclusive. The distinction was explained to them and they were asked to work independently to determine whether such terms could be collected with a fair degree of agreement. The trial showed that, while there were marginal differences for almost every worker, the lists which they brought back contained almost the same terms. This percentage of terms on which they agreed averaged so high that it offered a sufficient degree of objectivity to satisfy the requirements of the new-type test technique.

Encouraged by this degree of success, it was decided to subject the whole range of public school instruction in the social sciences to similar analysis. The range of public instruction was interpreted to extend from the first grade through the junior college, or sophomore year of college. It might be assumed that a comprehensive analysis of the material used in college courses in the social sciences would yield every term expressing social relationship to be found in the earlier years. Nevertheless, the texts and more common reference materials used in the elementary and secondary schools were also analyzed. As an additional precaution, an analysis of news columns of typical newspapers was also made. This seemed desirable in view of the widespread belief that there was insufficient relation between instruction and the problems of daily life. A comparison of functional facts of social science as found in the

TESTS AND MEASUREMENTS

newspapers and as found in the school books promised to reveal any such disparity.¹

LIMITATION OF THE LIST OF FUNCTIONAL FACTS

It soon became apparent, as the compilation of the list proceeded, that the list would grow so large as to become unwieldy. If the list was to be submitted to consensus appraisal, it was imperative that some limitation be put on it. But on what basis? The solution of this problem proved to be a difficult one, but it also led to a number of interesting and perhaps valuable observations. The successive limitations required were usually determined after a conference with two, three, or more persons. At times the whole working staff of the Investigation participated. The aim throughout was to obtain a list² by means of which as much understanding of the social sciences as possible could be determined. The principles used in limiting the list of functional facts follow:

1. *Requirements of New-Type Test Technique.* The use of the new-type test technique, which is intended to sample as much as possible in as short a test as possible, imposed its own limitations. Ideas of social relationship which could not be expressed in a single term, or, at the most, in a brief phrase of three or four words were automatically eliminated. This principle of elimination had the practical advantage of lending itself to fairly exact application by any number of workers.

Subject-matter specialists tended to resent this purely mechanical restriction, because it seemed to eliminate all the more complex ideas which could only be expressed in longer statements. Some of them thought that any tests constructed on so narrow a basis would yield only the beginnings of understanding in the field and leave the essential problem of genuine advance in understanding virtually untouched. On the other hand, all efforts to reduce

¹Dr. Edith E. Ware, formerly Dean of Russell Sage College for Women, devoted much of her time as special investigator to the analysis of college courses and newspapers. This work was further supplemented and checked by the late Dr. Frances E. Baldwin during the year in which she served as special investigator for the Investigation. Additional checks were afforded by Dr. E. E. Ellis of the University of Missouri and Miss Theresa Little of the University of Wisconsin. The list as compiled and checked by these workers appears in the Appendix.

²See complete list included in Appendix.

LIMITATION OF THE LIST

the larger ideas to manageable units for other than essay type questions had thus far failed. Analyses of textbooks and courses of study for either topics treated or the amount of space devoted to each topic had proved fruitless. Nearly a hundred such studies, chiefly unpublished master's theses, were consulted in vain. The only manageable material which they presented consisted of specific bits of information or exact verbal definitions. Not much more encouraging was the direct attack upon the larger ideas or "basic generalizations" made by Billings. The 888 generalizations contained in his list offered very few that could be used. Many of them were stated in terms which most social scientists could not accept without important qualifications. Most of those that could be accepted were so broad as to be practically meaningless. They did not, in the form in which they were stated, lend themselves to the desired economy of test-construction. The course suggested, by our advisor, T. L. Kelley, seemed, therefore, the only alternative.

Though undertaken reluctantly, this approach to the problem developed certain unexpected advantages. The first of these to appear was the realization that the method adopted (yielded simple units of ideas, each unit capable of identification, if not always of precise definition. The larger ideas, or basic generalizations, were compounded of these unitary ideas. Tests constructed out of the unitary ideas would, therefore, throw important light upon the pupil's mastery of the larger ideas). Another advantage appeared in the fact that the list contained unitary ideas which were learned only as parts of larger ideas. The pupil's mastery of these more remote units would accordingly imply a much larger knowledge and understanding.) To this extent, the procedure, which seemed at first too mechanical, yielded its compensations. It was hoped that perhaps the final results of such tests might afford even greater satisfaction.

2. *Terms Expressing Relations Among People.* It was early decided to distinguish between terms designating relations among people and those designating relations of people to the material world. To include both would have made the final list too large for

practical purposes. In choosing to omit the terms designating relations of people to the material world, undeniably "functional" though they were, the staff was moved by two considerations. Besides the size of the list, it was felt that for purposes of test economy knowledge of relations among people would strike more nearly at the essence of the social sciences. It might be assumed that pupils who showed mastery of these less tangible ideas would have less difficulty in learning ideas which focused on material objects. Such tests would more nearly reveal both training and ability. Whether this consideration can be entirely justified or not, its effect was to eliminate many "functional" facts used in history, economics, vocational courses, and geography, leaving political science and sociology relatively undisturbed. The line of demarcation could not always be sharply drawn, especially when the term emphasized the relationship among people, with reference to something in the material world. In general such terms were retained, though consistency proved difficult and some errors have undoubtedly crept in, both of omission and inclusion.

3. *Elimination of the Simplest, Most Obvious, and Most Frequent Social Relationships.* It seemed proper to exclude functional facts that every normal child might be expected to know satisfactorily before he reached the fourth grade. Such terms of social relationship as mother, father, sister, and brother did not promise to be very serviceable in a testing program. These terms and others almost as obvious occurred so frequently that their elimination promised a considerable saving of time and energy. There was a marginal fringe of judgment here, as on several other principles of elimination, with the probability that errors have occurred on both sides of the margin.

4. *Figurative Language and Synonyms.*¹ One of the most per-

¹This topic suggests another equally important, the lexicographer's semantics. Horn, in connection with his studies in spelling and reading, had occasion to consider the different meanings in which very simple words were used. He found no less than eighty-six different meanings attached to the word "get." The extremes to which society can subject a single and simple word is illustrated in the term "fast," which regularly designates both swift motion and total lack of motion. It is possible that the more common words are subjected to more meanings than are the more learned words, though it would be difficult, indeed, to find in the *New English*

LIMITATION OF THE LIST

plexing problems to arise as the list grew was that of dealing with figurative language. The same idea of relationship appeared in a variety of forms. This was more evident in newspapers than in books, more frequent in reference books than in textbooks, but not entirely absent from any of them. In most cases the variety arose clearly from a desire to avoid frequent repetition of the same word, however superior that word might be. The outcome of an athletic contest as described by a sports writer in a newspaper may appear as a "victory," or "defeat," or "rout," or "shambles," or "massacre," or "Waterloo," or nearly all of these in the same article. The range of synonyms in the more sober school books is seldom so extravagant, but both synonyms and figurative terms are used. More or less exact synonyms could also be classified as colloquialisms, localisms, terms of sophistication, and standard terms. Preference in the choice of synonyms was accorded to the standard term as having universal currency and more precise meaning. Only such of the others as had acquired a very wide circulation and common meaning were included. Here again judgments were frequently difficult and arbitrary, with the probability of error.

5. *Different Forms of the Same Word.* Another source of confusion, but one which promised some prospect of elimination, was that afforded by different forms of the same word. Ideas of the same relationship were found to appear in a definite series of forms, usually, though not always, variations of the same word. The simple idea of relationship appeared in noun form. When in process, it appeared in verb form. When used in a comparative sense, it took on an adjective form. When used to designate a functional use of the idea, it appeared in adverbial form. When the idea had become sufficiently common or widespread to acquire a permanent organization to support it and a fixed location or headquarters, it appeared in what might be called an institutional form, usually indicated by the suffix "-tion" or "-at." When the idea was a matter of controversy, this was often indicated by the prefix "anti-" or "pro-" or

Dictionary words of only a single meaning. It would be most difficult to find such terms in the social science field. The association of a single meaning with a given word must, therefore, be regarded as an ideal rather than an assumption.

TESTS AND MEASUREMENTS

the suffix “-ist.” When the idea had come to have fixed associations with a more or less definite group of other ideas and attributes, this was indicated by the suffix “-ism.” Sometimes the word “-minded” or “-conscious” was used as suffix to indicate what might be termed the ideological form of the word.

It appeared that the list could be materially reduced, if only one form of the word were included. The simple form of the idea, the noun form, was given preference unless the institutional and ideological forms represented so much modification of the original as to constitute virtually separate ideas. Thus, society, social, and socialism had all to be included. So, too, brother, brotherhood, and fraternity. It was not always easy to determine when the degree of difference in meaning had reached proportions necessary to include another word as a separate term in the list. There are undoubtedly errors on both sides of the line.

6. *Obsolete and Foreign Terms.* It seemed fair to omit terms that once served society, but were no longer used to express the idea. Terms in foreign languages sometimes used in English for an idea of relationship adequately represented by an English term might be characterized as sophisticated and omitted on that basis. Terms rarely used could also be omitted, if the idea was commonly expressed in another term. At least it seemed fair to exclude them, in order to keep the list down to manageable size.

The effort to do this, however, also occasioned its difficulties. There was little difficulty in eliminating such a word as “fief.” On the other hand, what was to be done with a term like “slavery?” The United States had formally abolished slavery in 1863. But the League of Nations found millions of people still in a state of slavery as late as 1930. Obviously the term “slavery” was not obsolete. Another source of difficulty was found in the fact that many fairly common ideas were expressed by different terms in successive generations or even more frequently. The margin of choice was much wider in applying this principle than in most of the others and the possibility of error consequently much greater.

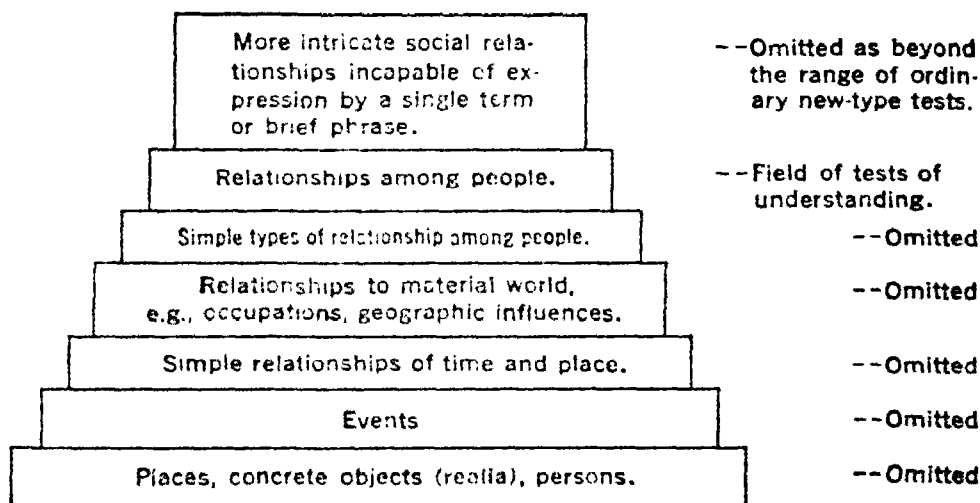
7. *Hierarchy of Officials.* Another source of perplexity came from the variations in title used to describe the hierarchy of officials in

SIGNIFICANCE OF THE FINAL LIST

the greater variety of organizations, social, economic, and political. The differences between the offices themselves, as ideas of social relationship, were too slight to justify the inclusion of all the varied forms by which an essentially similar office was entitled in different lands or social groups or industrial groups. It was decided, therefore, to retain in each case a complete hierarchy of titles as the type, and to add only such additional titles as special considerations made necessary. In this case, too, the margin was very wide, and, in fact, necessitated selection between two margins, with corresponding difficulty and probability of error.

SIGNIFICANCE OF THE FINAL LIST

The relationship of the final list to the whole body of social science materials may be represented in diagram form:



The list as finally compiled has a qualitative rather than a quantitative significance. No direct effort was made to determine frequency of occurrence. This entered into calculation incidentally as a basis for the elimination of rare, unusual, and obsolete terms. It may, therefore, be assumed that all the terms in the list occur frequently enough to be a potential charge upon learning in the social sciences.

The defect in qualitative range suggested by the application of the mechanical requirement of no more than a single phrase proved

to be less serious than at first supposed. The larger ideas in their ramifications include many terms which occur only in the matrix of those larger ideas. Many of these were caught in the list, and their inclusion in a test, therefore, might imply, if not fully prove, the understanding of the larger idea.

Though still lacking the more elaborate and complex ideas expressed in sentences and paragraphs, which may represent the most advanced knowledge, the list affords a fair sample of the range of ideas taught in the social sciences through the sophomore year in college. It actually touches a number of ideas not usually acquired until later in college life.

A student who showed a satisfactory knowledge and understanding of the functional facts contained in this list might be presumed (1) to have studied the social sciences through the junior college, (2) to have an understanding of social ideas beyond the average adult, (3) to have capacity for advanced study of the social sciences.

The inclusion of the whole range of courses offered in the social science field in the grades from the fourth through the junior college—history of all periods, civics of several kinds, economics in several forms, sociology and geography, and the contents of a year's run of several current newspapers—afforded a wide and probably unusual view of the growth of social ideas. Some of the observations made by the several workers engaged in the task may be of value, therefore, in suggesting hypotheses as to the learning process in this field.¹

1. The distinction between "functional" and "inert" or "unique" facts in social science seems to be valid in the sense that different workers employing this distinction independently arrived at lists that were almost identical.

2. Inert or unique facts must be qualified as ranging from literally inert to a considerable degree of functional value, or, rather, inert facts can be best described as incomplete functional facts.

This observation may require some explanation. Social ideas are

¹For a fuller discussion of this point, see pp. 45-53, 109-119.

learned in many ways—outside the classroom as well as within. When learned outside the classroom, the learner tends to associate the idea learned with some person, place, or material object present when the idea dawned upon him, whether there was any such essential connection or not. He refers henceforth to the idea in terms of the object, place, or person with which he has associated it. If the idea is one of common occurrence and the association used appeals to others as preferable to the term they then use, or if it supplies a name for an earlier idea, it gains general currency and is well on its way to the status of a functional fact. It may be said to have reached that status when the proper noun which is usually employed at the outset becomes a common noun.

Thus the idea of wanton destruction became associated in the minds of medieval churchmen with the sack of Rome in 455 by the Vandals. The association acquired a permanence and popularity until today "vandal" and "vandalism" are common nouns indicating that idea. Thus, too, the idea of mass production, with its complex of social connotation, has become identified in European literature with Henry Ford, and the term "fordisme" and "fordismus" have already become functional terms in Europe. Many politicians today use the term "gerrymander" unconscious of the fact that it is derived from the name of an individual associated with that type of political redistricting. Wall Street has become a synonym for financial speculation, as Broadway has for theatricals. They may be cited as examples of proper nouns well on their way toward the status of functional terms. Tammany Hall as a synonym for machine politics has already moved far in that direction. "Tammanyism" needs only to alter the capital letter to be said to have attained this status.

3. Functional facts, or terms describing very common social relations, are in constant danger of being supplanted by other terms. Doubtless many factors enter into the explanation of this phenomenon. The fact that such terms represent extra-school relations; the desire to vary the monotony of using the same term for an idea which so frequently recurs; an unconscious feeling of incongruity about applying a "book" word to a life situation, as

well as the tendency to describe the situation in terms of association, all play a part. Every generation, it seems, adopts its own terminology to describe courtship and its associated ideas. A lexicographer could be kept busy listing the various expressions that have been devised to describe direct money grants by Congress alone.

4. There seems to be an increasing order of difficulty in the various forms of an idea as expressed by different forms of the word. Only the simple idea in noun form registers a high degree of accuracy and precision in use.¹ Next in simplicity to the noun apparently comes the verb form. The adjective follows, but the adverb is much less used. The institutional form represents a fairly high degree of learning, while the ideological form is seldom used accurately. The basis for this observation is the frequency and varying connotations with which these forms were found both in newspapers and in textbooks.

5. Only a relatively small proportion of the functional facts of social science may be regarded as definite and permanent. Those terms which are used only by the scholar or the advanced student have the best chance of retaining a precise meaning. Those which serve the whole population seem to be in constant danger of modification. This observation is most strikingly confirmed by the fate of terms coined by scholars and usurped by the general public, as, for instance, those connected with psycho-analysis. It is likewise illustrated by the recent efforts of scholars to state some of their abstract findings in simple language.

DIFFICULTIES ENCOUNTERED IN EVALUATION OF ITEMS

It soon became apparent that the preparation of the comprehensive list covering all social science work through the junior college years would require a large amount of time. There was also strong reason to believe that the further step in the preparation of the tests would require as much, if not more, time. Under the circum-

¹The relative difficulty of the simpler forms is not so strongly marked. Greater vividness or concreteness or more frequent use may in some cases reveal a verb as more easily learned than the noun and, in fewer instances, even the adjective and adverb. In general, however, the order of difficulty appears as stated.

DIFFICULTIES ENCOUNTERED

stances, it seemed wise to use the original list compiled from text and obvious reference books used between grades four and twelve as the basis for preliminary test work in understanding. This list of some three thousand terms was sent to thirty college teachers of recognized scholarship and extended teaching experience with beginning classes in the social sciences in all sections of the United States and in all types of educational institutions.

These judges were asked to grade each item on the following bases—whether they expected the graduates of high schools enrolled in their courses to show (1) complete understanding of the term, (2) some understanding, (3) no understanding. The judges were also asked to suggest terms not on the list which they thought should be included. For convenience in checking, the list was prepared in alphabetical, rather than in logical, form.

These instructions, as indeed would any instructions, invited differences in evaluation. The instructors inevitably judged a term on the basis of its usefulness in the courses they were conducting with freshmen classes. Some terms essential to a course in Modern History might not prove necessary in teaching American Government. Similar differences in evaluation were inevitable in other fields of social science. Hence terms absolutely essential in one instance might be designated as of no value in another.

Later, when the lists in special fields of social science were prepared, similar difficulties of evaluation were discovered. Classified lists in economics, political science, and sociology were submitted to experts in these fields. This procedure proved unsatisfactory because there were marked differences in classification followed by different specialists in these fields. These differences were so marked that an expert accustomed to one form of classification found it practically impossible to evaluate the items in a different scheme of classification. The method had to be abandoned for the more objective and less controversial alphabetical arrangement. Even this, however, presented difficulties. Specialists, especially in the field of sociology, were found to have a different nomenclature for ideas which laymen might regard as practically the same. This diversity was due to differences in point of approach. Judgments

TESTS AND MEASUREMENTS

of the value or importance of terms which sociologists do not generally employ varied even more than did the judgments of the college teachers concerned with freshmen courses. While sociology has been cited as presenting this difficulty, it proved scarcely less true in the fields of economics and political science.

Another difficulty encountered in the evaluation of the list appeared most pronounced in dealing with political science. While there was much less difficulty about variety of terminology in this field than in either sociology or even economics, the mechanical limitation of confining the ideas to single words or brief phrases proved an almost insurmountable obstacle. The specialist in political science found himself halted by wondering in what connection the term was to be used. It took on a different value as it appeared in different connections. Most political scientists, it became evident, would face such a task with a sense of complete frustration. Only a very few political scientists, and those most nearly connected with teaching in the lower schools, could be persuaded to disregard their scholarly demand of qualifying connections even to attempt the task of evaluation. This difficulty was most pronounced in the field of political science, but it also occurred in economics and sociology.

Still another difficulty arose from the size of the task itself. Despite all eliminations, the list was still a formidable one. The list for the grades four to twelve contained nearly three thousand terms, while the combined list for the social sciences in the junior college grades was even longer. An accidental omission of a page of the mimeographed copy of the first list revealed this difficulty. The task of judging each of three thousand items involved not merely the evaluation of the separate item but its evaluation with reference (1) to the demands of the whole range of a year's work, (2) to the presence or absence of other terms in the list which more nearly satisfied those demands, (3) to a possible omission which might more nearly satisfy the judge's needs. Since each of the judges was abundantly occupied with his own immediate work, weighing the value of each of these terms in this circumspect and comprehensive manner was virtually impossible. This was fully

DIFFICULTIES MET

proved by the fact that only one of the thirty judges noticed the omission of a page containing terms beginning with the letter "s." The one person to notice the omission was the only one of the judges who indicated any omissions. The evidence of fatigue as well as difficulty in finding time to finish the task was revealed by the fact that in several cases the terms on the last page or two were not checked.

In view of these difficulties encountered in the evaluation of the terms, the application of a rigid statistical technique to give a consensus evaluation or weighting to each term in the list would seem rather the compounding than the obliteration of error by the combination of judgments. This was done, however, to satisfy the demands of objective technique, and, after the weightings were statistically established, the terms were arranged in quartiles according to their weightings for importance. Inspection of the results seemed to indicate that the more rarely used and the most difficult were in the upper quartile, while the most essential appeared to be those easiest or most commonly used. Following this inspection as a clue, it was decided to construct tests chiefly from the two middle quarters.

DIFFICULTIES MET IN CONSTRUCTION OF TESTS

This weighted list of sub-collegiate functional facts made it possible to begin the work of test construction in accord with the principles established for this work. Even in this stage of the work Dr. Kelley insisted upon a combination of subject-matter specialists and psychological technicians. Finding persons capable of constructing tests proved more difficult than had been anticipated. Finally Dr. Edgar B. Wesley, head of the department of social studies in University City High School of University City, Missouri, Miss Mary Gold, head of a similar department in the University of Minnesota High School, and Miss Mary G. Kelty, formerly in charge of this department in the Oshkosh Normal School, and Dr. Luella C. Pressey of the department of psychology of Ohio State University were engaged to construct these tests. All had demonstrated interest and ability in this work. Dr. Pressey

TESTS AND MEASUREMENTS

approached the problem from the more objective standpoint of a psychologist long interested in vocabulary studies, especially her own covering elementary and junior high school levels, while the others represented primarily the interest of the subject-matter teacher. Dr. Wesley was primarily interested in the senior high school, Miss Gold in the junior high school, and Miss Kelty in the intermediate grades of the elementary school.

This work of test construction also proved to be very time consuming. The detailed account of the construction of these tests is told elsewhere.¹ It will be sufficient here to summarize the experience of those engaged in the making of the tests. Hundreds of questions were prepared. A variety of forms were used, the true-false, the completion, the matching, and the multiple-choice forms being most favored. The questions were tried out and the results statistically analyzed. After a series of trials, the multiple-choice form of question was selected as best for testing understanding.

The completion type of question, which gave the student wide freedom in supplying missing terms, phrases or clauses or definitions and illustrations of ideas or examples of various kinds, proved most revealing to experienced teachers. The choice of answer in many instances revealed how far the student was from the desired understanding of the idea involved. Unfortunately, however, the variety of answers was so great and represented so many degrees of understanding that only an expert in the knowledge of the subject would be able to grade such answers satisfactorily. It was found impossible to construct a scoring key for such questions which an ordinary clerk could employ. Despite the high diagnostic value of this type of question, the fact that it required experts and consumed more time in scoring made it too costly for our purposes. It was abandoned with regret.

The true-false type of question, while apparently easy to construct and easiest to score, presented other difficulties. In the first place the ease of construction was apparent, not real. Very few questions in social science can be answered with a categorical yes or no, right or wrong. To construct questions which could be so

¹See chapter III.

DIFFICULTIES MET

answered proved to be most difficult. Secondly, these questions lent themselves too readily to pure chance performance. Incorrect answers might prove ignorance, but correct answers might represent mere chance. The statistical procedure of multiplying the number of questions so as to overcome the factor of chance seemed too cumbersome for general use. This type of question was retained by two workers--by one who had found through years of trial a few questions so worded as to catch the interest and challenge the knowledge of junior high school pupils, thus eliminating the chance element, and by another who, by using diagrams and pictorial illustrations equally challenging, also overcame the chance element. Both of these workers, however, were extraordinarily able and had arrived at these questions through years of experience. They had managed to find questions which pupils could, but would not, answer by mere guessing. Under ordinary conditions, however, this type of question proved least satisfactory.

The matching type of question in which a number of related or comparable items are placed in two columns, the student to place before the items in one column the number of the related or comparable item in the other, proved somewhat better than the true-false or completion types. The question involved some degree of relationship, which was desirable, and it could be scored almost as easily as the true-false form. However, it afforded considerable opportunity for guess-work--too much if the columns were of equal length. Guess-work was cut down somewhat by making one column longer than the other, thus including some items that did not fit. There was a further difficulty in that the mechanical labor of answering such questions was greatly increased by lengthening the columns. If the columns were very long, the time required to answer the question was too great, the eye-strain too severe; if they were short, the opportunity for mere guess-work was too great. Though useful, this type of question was not as satisfactory as the four- or five-option multiple-choice type.

The multiple-choice type permitted the use of phrases as well as of single terms. It afforded opportunity to test relationship as well as recall. By the use of four or five options the chance factor was

TESTS AND MEASUREMENTS

cut down to more controllable proportions. The options in the test could be gauged to a relatively fine discrimination of the student's knowledge. Each item was sufficiently brief and compact to be answered without unusual eye-strain, and the answer could be scored quickly and accurately with a scoring key. Both Mr. Wesley and Miss Kelty finally used this form alone, as did also Mrs. Pressey, who was working independently.

In the course of determining the most efficient test form, it became increasingly apparent that the value of the questions depends less upon the form than upon the content. The mechanical advantages of the new-type question may easily be completely vitiated by careless selection of items. In fact a casual, careless, thoughtless test in the new-type form may prove not only unsatisfactory as a test but may impair or defeat good instruction. This danger has been generally overlooked by test-makers. A careless essay-type test at least affords the student a chance to display his knowledge and ability by the way in which he develops his answers. A new-type test, on the other hand, limits the student to what is set before him. If that is bad, he has no recourse. The advantages of the new-type test are too largely mechanical. The value of the test as a measure of achievement depends upon the matter cast in that form as well as upon how it is cast.

The wording of the questions proved even more exacting than the choice of forms. The vocabulary of pupils at different grade levels proved to be a conditioning factor of the utmost importance. Wording that seemed awkward or even childish to college students was frequently necessary to test the degree of understanding of the idea possessed by the pupil in the elementary or even junior high school. Not infrequently questions were condemned by college teachers, only to have the statistical analysis of results in the lower grades seem to prove that the questions were actually testing the pupils of the lower grades with apparent accuracy. The more precise and universal vocabulary of adult levels was not familiar to the pupils of the earlier grades. At any rate, it afforded an uncertain approach to their understanding of the ideas involved.

The range of the options used in the multiple-choice form proved

also to be a matter of concern and importance. If only one of four or five options bore on the question, while the others were very remote, the pupil's selection of the correct option did not necessarily prove any real understanding. Often only the 'vague' impression of association between the question and the correct option was necessary to gain credit for understanding under such circumstances.

PROBLEMS RAISED IN THE SELECTION OF QUESTIONS

The questions as drawn up by Wesley, Gold, Kelty, and Pressey were submitted to the inspection of at least three experts in the knowledge of the subject. This inspection resulted in the elimination of some questions because of errors of fact, impression, or statement. The maker of the test was given the opportunity to review these judgments and in some cases, notably in the case of questions made out for elementary and junior high school grades, the test makers insisted upon retaining some of the objectionable questions. It was their contention that the questions, though awkwardly worded and perhaps objectionable on the adult level, afforded a truer approach to the degree of understanding possessed by pupils at the lower levels. Fortunately several of these questions were retained for the actual trial of the tests.

When the tests had passed this inspection and had been scrutinized by Dr. Kelley for technical form, they were tried out under typical school conditions. As indicated in the detailed account of the Wesley test,¹ nearly every one of the tests went through a series of trials. For the final trials the tests were given to pupils in different parts of the country, and, wherever possible, in a variety of school systems. The tests were sent to Dr. Kelley under whose direction they were scored and the results submitted to statistical analysis. Statistical profiles were worked out for each item and these were then submitted to the advisory Committee on Tests with Dr. Kelley's recommendations for each item. The effect of this process of checking and re-checking the tests was to cut the original list of approximately four hundred items in the Wesley test to about one hundred and sixty.

¹See pp. 219-227.

The use of a rigid statistical technique in selecting the items to be included in the final test was not accepted without misgiving by the subject-matter specialists. They saw some questions which to them seemed among the best testing items dropped, while others relatively poor were retained. Items were selected on the basis of progressive performance grade by grade, the item showing the steady improvement apparently being regarded as the best. This resulted in the elimination of some questions which most teachers would regard as unusually good ones and the inclusion of others which seemed relatively poor questions. Furthermore, since the statistics were obtained from a wide range of pupils and school systems, what relation did they bear to instruction? The school systems were offering different courses in the same grades. The classes, especially in the later grades, included pupils who had taken no courses in social science for two, three, or even four previous years as well as others who had taken courses every year. The question whether a test constructed on this statistical basis under such conditions would measure anything but the growing maturity of the pupils was very seriously raised. Did not this procedure tend to wipe out all evidence of instruction? If it did, of what possible value could it be in the measurement of instruction?

The only qualifications in extenuation of this procedure which the teacher could see at this time were: (1) the fact that the questions used were all drawn from social science subjects, and even if the test measured growth alone, it would be measuring social growth; (2) that all the pupils tested in the trials had received some formal instruction in social science, and even those who had omitted it during most of their high school years were receiving informal instruction from other sources during that time; (3) that the application of the final test, apparently one of social growth alone, to instruction might show that different courses or different teaching procedures might vary in their contribution to that social growth. If so, such differences might have a far greater significance than if gleaned from a more definitely instructional test, whose value still remained to be determined. But these were all questions that would require later answer. With these misgivings, but with a determina-

EXPERIMENT ON ACQUISITION

tion to follow out the dictates of objective measurement, nevertheless, the final tests were drawn up on that prescribed basis.

SUPPLEMENTARY STUDIES IN TESTING UNDERSTANDING

EXPERIMENT ON ACQUISITION OF "INERT" AND "FUNCTIONAL" FACTS BY HIGH SCHOOL STUDENTS

Fortunately, other workers were developing tests during these years, and the Chairman of the Commission had the assistance of a seminar, many of whose members were particularly interested in this problem. With their help many of these questions were attacked and some suggestions toward their solution, if not the final solutions, are possible. Four projects undertaken by members of this seminar at the University of Minnesota and the work of Lindquist and Anderson at the University of Iowa may be taken as examples of this supplementary work.

The first project was an experiment to determine the relative learning of "inert" or "unique" and "functional" facts. This has been reported in full elsewhere¹ and may be summarized here. Modern history was chosen as the subject-matter field, partly because it is unusually rich in "inert" or "unique" material, predominantly political, and partly because it is generally regarded as difficult. A course in modern history is offered in the second year, tenth grade, of high school and again in the first year of the university. Two groups of seniors in the University of Minnesota High School who had not elected modern history were used for the purposes of the experiment. Their performance was checked by comparison with tenth grade classes in the University High School and in one of the Minneapolis high schools, as well as with a section of the university class in modern history. The first experimental group was given a week in which to work up the whole field of modern history. Though given the opportunity to obtain assistance from experienced teachers, they chose, after the first two days, to do all the work themselves. They had available three of the most widely

¹Krey, A. C., and Wesley, E. B., "Does the New-Type Test Measure Results," *Historical Outlook*, XXIII (January, 1932) p. 7.

TESTS AND MEASUREMENTS

used textbooks in modern history, including the one used by the pupils in the tenth grade, and, for further reference, one or two textbooks used in college. The same procedure was followed with the second experimental group of University High School Seniors three years later.

For the first test, a group of experienced teachers enrolled in the seminar, went through the high school textbooks and constructed the examination on the principle of covering the whole field, using important items rather than obscure details and avoiding questions which required a new or original rearrangement of the information. The new-type forms of questions were used in the test. The test was submitted to a number of experienced teachers for criticism before it was applied. For the second test, Dr. Wesley, assisted by Miss Mary Gold and Miss Dorothy Bovee, selected "functional" facts which were especially emphasized in modern history, terms dealing with international relations, comparative governmental and social institutions. They used the multiple-choice form of the new-type examination. Both tests were applied simultaneously to the university group and to the two high school groups. The scores follow:

TEST OF "FUNCTIONAL" FACTS OR UNDERSTANDING

Score	Experimental Group	Selected High School	High School	College
90.....		x	x
85.....		x	xx	xx
80.....		xx	x
75.....		xxxxx	xx	xxxxxxxx
70.....	x	x	xx	xx
65.....	xx	xx	xxxx	xxxxx
60.....		xx	xxxx	xxx
55.....	xx	xx	xxx
50.....		xxxx	xxxxxxxxxxxx
45.....	x	xxxxxxxxxxx
40.....		xx	xxxxxxxxx
35.....		xxxxxx
30.....	
25.....		xx
20.....		xxx
15.....		xx
10.....		x

EXPERIMENT ON ACQUISITION

TEST OF "INERT" FACTS OR INFORMATION

Score	Experimental Group	Selected High School	High School	College
155		X
150	
145	
140	
135		X
130	
125	
120	X*	XX
115		XXXXX
110		X
105		XXXX
100	
95	X	X	XXXXX
90		X
85	X	X	XXX
80	X	XXX	XXX	XXXX
75	X	XXXX	XXX	X
70		XX	XX
65		X	XX
60		X	XX	X
55		XXX	X
50		X	XXXXX	X
45		XXXXX
40		XXXXX
35		X
30		XX
25		X
20		XX

*After this pupil completed the test it was learned that he had had a Modern History course in another school.

Interpretation of the Results. Since the number of pupils was small, only tentative inferences may be drawn from the experiment. These, however, may be listed as follows:

(1) The experiment lends support to the conclusion of the special investigators compiling the lists of functional facts, that (there is an important difference between inert, or unique, and functional facts in social science)

(2) (That inert, or unique, facts are much more readily learned than functional facts)

(3) (That an examination or test constructed chiefly of inert, or unique, facts does not measure instruction in the subject. Such a test can be satisfied by cramming)

TESTS AND MEASUREMENTS

(4) That an examination or test constructed chiefly of applied or differentiated functional facts seems to require something beyond cramming and reflects sustained instruction or reflection, or both.

The second project to be considered here was a study of the results obtained on the Wesley-Gold test by all the pupils of a junior-senior high school. The classes in these six grades averaged about 100 pupils per grade. Careful I. Q. ratings had been established for every pupil in the school. While nearly all of the seventh and eighth grade pupils were taking some course in social science, a number of ninth grade pupils, nearly two-thirds of the tenth grade, one-third of the eleventh, and about half of the twelfth were not taking work in this field at the time of the test. This project thus afforded some opportunity to compare the results on the test of those who were and those who were not enrolled in social science courses in each grade, as well as to compare the results with the I. Q. rating of the pupils.

As expected, there was a definite improvement in the scores registered by each successive grade up to the twelfth. Pupils not enrolled in a social science course at the time of the test showed improvement, as also did those enrolled. A comparison of the test scores of those who were and those who were not taking the concurrent course in social science revealed several points of difference. When the relative rank of the pupils in each class as determined by their I. Q. ratings was used as the standard, those who were taking concurrent work in social science improved their rank order on the Wesley-Gold test, while the others fell below their I. Q. rank in every grade except the twelfth. The margin of improvement was greatest in those years in which some history course was taken, the improvement being equally marked whether the field was ancient, modern, or American history. Again, as in other trials of the test, pupils in the lower quarter of the I. Q. rating registered the greatest improvement in those grades in which history was the concurrent social science study.

The fact that at the twelfth grade there was no difference between those enrolled in the social science course and those who

RELATIVE CONTRIBUTION

were not was puzzling. This might seem to indicate that the test measured a social growth which had been attained at that age without reference to instruction. Or it might mean that the test had reached its ceiling by that time. Investigation, however, revealed that in this particular high school all pupils were required to take either the eleventh grade course in American history or the twelfth grade course in social science. The comparison of results on the test at the twelfth grade was, therefore, a comparison of those who were taking social science concurrently with those who had had American history the year before.

EFFORT TO DETERMINE RELATIVE CONTRIBUTION OF DIFFERENT SUBJECTS TO KNOWLEDGE OF FUNCTIONAL FACTS

The third project undertaken by members of the seminar was the trial of the Wesley-Gold test under ordinary school conditions.

	Courses Pursued in Intervening Grade	Geography	American History	Community Civics	Modern History	American History	Social Science
Score	6A	7A	8A	9A	10A	11A	12A
95	2	1	1
90	1	5	12	7
85	11	22	24
80	5	24	20	24
75	..	1	4	14	34	68	39
70	..	2	3	31	36	61	54
65	..	3	19	41	48	46	36
60	1	11	27	84	37	37	26
55	3	17	52	108	25	22	7
50	10	31	44	121	16	7	8
45	23	39	46	99	7	3	1
40	41	84	40	81	2	4	..
35	46	98	21	45	1	1	..
30	50	71	16	18
25	40	47	5	9
20	14	23	1	..
15	6	7	..	1
10	2	2
Number of pupils...	236	436	277	658	248	305	227
Median...	35.7	38.5	51.2	53.0	68.7	72.6	73.3
Differences.....	+2.8	+12.7	+1.8	+15.7	+3.9

TESTS AND MEASUREMENTS

Miss Prudence Cutright and Mr. Walter Anderson, both members of the Research Bureau of the Minneapolis Public Schools, tried this test on a sampling of at least two hundred pupils of every grade from the sixth to the twelfth, inclusive. The test was administered near the close of the school year, thus affording some evidence on what influence, if any, the current work in social science might have upon the scores. (See table on page 29.)

The small gain in the 9th grade, in which Community Civics was the course offered, was disturbing. More of the terms used in the test, the ideas involved, were mentioned in this course than in any other except the 12th grade course in social science. Yet every other grade except the last showed a greater gain than did the 9th grade. The most significant gains were made in the American history course of the 8th grade and the modern history course of the 10th grade. Fearing that the showing of the 9th grade might have been due to some disturbance caused by the fact that this grade marked the end of the junior high school, the test was applied to 9th grades working under other conditions. A neighboring rural school system following the older 8-4 plan of organization tried the test. Here the 9th grade showed a marked improvement, but the course offered in that year was ancient history, not community civics. The University High School offered another check. It, too, was organized like the Minneapolis schools on the 6-3-3 plan, but, unlike that system, it offered ancient history in the 9th grade. Here again there was a marked gain. The explanation of the difference would, therefore, seem to lie in the course itself, or in the way in which it was conducted.

Another effort to check these results was made with the help of the St. Paul school system. Here the course in community civics was offered through both 8th and 9th grades, two years being devoted to the work instead of one. A sampling of about one hundred pupils for each grade from the 7th to the 12th revealed a somewhat greater gain for the work in community civics, though even there the gain registered by the history courses was greater, American, ancient, and modern history indiscriminately showing the greater gain.

RELATIVE CONTRIBUTION

In order to check the showing of the 9th grade still further, Cutright and Anderson constructed a test based directly on the work offered in the course as conducted. They included in their test a number of questions prepared by the author of the textbook. Though making an effort to keep their test one of understanding rather than of information recall, they included a number of definitions as offered by the text and derived their functional fact items from the information offered by the course. The test was administered to another sampling of pupils from the seven grades. The results follow:

Score	6A	7A	8A	9B	9A	10B	11B	11A	12A
72-74...						1	2	1	
69-71...					2	2	4		2
66-68...				1	7	5	2	2	2
63-65...				5	10	3	9	6	9
60-62...				13	9	17	3	8	20
57-59...			4	12	11	14	11	3	14
54-56...			3	25	22	19	7	7	8
51-53...		1	8	34	8	16	6	8	10
48-50...		1	7	29	12	31	7	10	9
45-47...		2	17	28	10	16	3	7	2
42-44...	1	2	11	21	9	20	3	6	6
39-41...	1	10	15	27	3	8	2	5	2
36-38...		13	12	18	2	6		4	1
33-35...	2	3	15	19	1	2			1
30-32...	6	13	19	9	1	3			
27-29...	6	13	11	4				1	
24-26...	4	2	10	2		2			
21-23...	7	5	4	1		1			
18-20...	4	2	2			1			
15-17...	6	4	1						
12-14...	2	1							
Total.....	39	72	139	249	107	167	59	68	86
Median.....	24.4	32.1	37.9	47.5	55.0	50.4	57.4	51.4	57.5
Gain from Grade to Grade.....	7.7	5.8		17.1	-4.6	2.4	-3.6	2.5	

These results showed the greatest gain in the 9th grade, the gain as registered at the end of the 9th grade being greater than the gain made during the previous two years. It remains to be determined what significance may be attached to the fact that, though all the pupils tested in the later grades had had this same course in the

TESTS AND MEASUREMENTS

9th grade, there was an actual loss in two of the grades and no appreciable gain during the following three years.

Another project which has some bearing upon the test of understanding was carried on by Cutright and Anderson, together with Mr. B. Lundquist, instructor in geography at the University High School. The Parker-Calkins Test of Geographical Understanding, Parts I and II, Form I was applied to a sampling of pupils in six grades, with the following results:

Score	6B	6A	7B	7A	8A	9A	10A	11A
135.....	1
130.....	1	1
125.....	1	2	2
120.....	4	5	2	2
115.....	1	1	1	3	8	6	7	7
110.....	3	4	13	3	27	11	9	21
105.....	7	9	15	14	25	30	17	20
100.....	14	11	28	30	39	37	15	19
95.....	20	25	23	42	40	42	17	27
90.....	33	43	51	49	46	38	13	23
85.....	58	33	47	45	38	31	6	20
80.....	47	31	43	54	26	24	10	7
75.....	30	28	27	22	8	10	2	4
70.....	26	12	20	15	6	1
65.....	10	8	5	7	5	1
60.....	8	6	2	6	3
55.....	4	3
50.....	1	1
45.....	1
40.....	..	1
Totals.....	262	213	277	294	275	236	100	153
Median.....	85.4	88.	89.4	89.4	95.7	96.5	100.7	99.2
Gain.....	2.6	1.4	0.0	6.3	.8	4.2	-1.5	..

These results indicated their own problem, the determination of the causes for the greater gain in the two history courses, grades 8 and 10, than in the course in geography, grade 7. Lundquist addressed himself to this problem. A check trial of the test in the University High School showed substantially the same result. Analysis of the items in the test and their behavior in the two trials afforded very little light. Analysis of the course in geography offered in the two school systems as well as the methods of procedure employed yielded no convincing explanation. One con-
 jec-

LEARNING OF ORGANIZED INFORMATION

ture, that the ideas acquired in the geography class demanded maturing and would not register until the following year, was entertained. This might explain the greater gain in the year following the work in geography. It did not, however, explain why there was no appreciable gain in the second year following, in the course in community civics, and why there was such a marked gain in the third year following, when modern history was offered. The only other inference suggested was that the geographical ideas acquired greater meaning as a result of the vicarious experience with those ideas which the narrative history courses supplied. No positive proved explanation was found.

LEARNING OF ORGANIZED INFORMATION

All of these projects carried some implications as to the ways in which children learn about society.¹ A closer analysis of some of the details yielded further implications. In the trial runs of the Wesley Test the term "sepoy" was used. This item was one of those which, according to the principles set forth earlier in this introduction, should have been eliminated. Unimportant in itself and easily learned, it seemed absurd to include it in the test. Though marked for omission as unessential, a typist's error permitted its survival. On statistical analysis of the trial results, it proved to be one of the most discriminating items of students' knowledge and interest in the social studies. Why?

While this question was still unsolved, some suggestions of an answer appeared from apparently unrelated quarters. About this time one of the investigators observed two small boys aged eight and ten engaged in building a locomotive. Their "working drawings," the result of much planning, were nailed to a tree and the materials for the enterprise, accumulated from various parts of the neighborhood, were piled up about it. With a four-wheeled wagon as the base and an empty oil drum as the superstructure, the project was well under way. An engine cab neatly constructed of broken laths was attached to the rear. Various cans were soldered

¹See pp. 14-16, 45-53, 109-119.

TESTS AND MEASUREMENTS

on the top of the oil drum. One was attached in front and a light was installed. A piece of pipe was attached as a chimney, and, after some days, the finished "locomotive" was moving about the neighborhood. The investigator recalled the sustained interest of these boys in locomotives. They had seen thousands of locomotives, had visited round-houses, clambered all over the old discarded engines and had even had the thrill of being shown over one of the most modern engines by a real engineer. He also recalled that these boys had been quite content to play with toy locomotives which represented little more than impressionistic shapes---had preferred them in years gone by to the more elaborate imitations. It was therefore doubly interesting to compare their ideas of a locomotive at the ages of eight and ten with their ideas of the same subject at four and six. The difference consisted primarily in the additional detail which they now demanded. Yet even at eight and ten, the elaborate motion gear without which a locomotive is not a locomotive at all was still missing. The "working drawings" provided for a cylinder and there was also a rudimentary indication of at least one driving rod, but practical difficulties had prevented their inclusion in the finished "locomotive."

In Miss Van Bibber's study of specific difficulties in the teaching of the social studies as indicated by teachers of those subjects, the Federal Reserve system was one of the topics most frequently mentioned. Teachers from all parts of the country and from all grades between the fourth and twelfth were included in her study. She was struck by the fact that in the case of some of the topics, notably the Federal Reserve system, the difficulty increased as the students advanced. Only a small proportion indicated it as difficult in the seventh grade, whereas the proportion of teachers reporting it as difficult seemed to increase with each grade, being greatest in the twelfth grade.

The connection between the "sepoy" item, the steam locomotive, and the Federal Reserve system is not very obvious, yet there may be a common element. The term "sepoy" gains prominence from the "Sepoy Mutiny," a detail of the history of British administration in India which is treated as one of the topics of modern

history. Only the student who had thoroughly studied his modern history and had been interested in problems of governmental administration would be likely to recall the episode. Hence the fine discrimination which the item exercised in the test. Growth in knowledge of the steam locomotive was likewise evidenced by the increased knowledge of detail. The increased difficulty in understanding the Federal Reserve system was doubtless due to the greater detail which was involved in its consideration in the upper grades. All three items represent highly organized information, and their understanding suggests the observation made by William James that "a person can see into a generalization only so far as his knowledge of its details extends."

A truer test of the pupil's understanding of such highly organized information is furnished by his knowledge of the relationship of a few scattered details to the whole than by a description of the whole. This is in accord with the popular observation that a person most advanced in understanding of a game is marked by his knowledge of its fine points, and games, too, are organized bodies of knowledge. No need to ask a person who knows a "hit and run" play whether he understands baseball, nor one who knows the "Vanderbilt convention" whether he understands the game of bridge. For the American pupil the "sepooy" item marked one of the finer points of governmental administration and modern history.

Dr. Francis E. Baldwin, in working over the returns on the expert appraisal of the functional facts in economics organized some of the topics in terms of increasing difficulty.¹ The topic of banks and banking may serve as an illustration.² Banking is an institution of modern society with which every adult has some acquaintance. A glance at this array of functional facts connected with banking might enable each individual to determine the degree of his understanding of the institution. The difference between the

¹Perhaps "increasing knowledge" would better describe the quality of the arrangement. Some of the minor items included among the more "difficult" are not in themselves difficult, but they would not usually come within the range of the student until he had learned much about banking.

²See following chart.

TESTS AND MEASUREMENTS

expert and the man in the street is determined at the right of the diagram, not at the left. It is unfortunate that the same terms, "bank" and "banking", summarize this knowledge, for it permits the man in the street to cherish the illusion that he has a full understanding of the institution. In fact, the less he knows about it, the

CHART SHOWING RATING OF TERMS USED IN MONEY AND BANKING

Bank	Assets	Check Currency Loan Silver dollar National bank Interest rate	Gold bullion Silver bullion Reserves Branch bank Monetary unit Silver certificate Gold certificate National bank system Gold standard Balance sheet Credit instruments	Mortgage loan Banker's bank Certified check Deposits as cur- rency Call loan Collateral loan National banknote Acceptance Chain bank Greenbacks	United States national debt Deprivation Federal Reserve bank Federal Reserve note Federal Reserve bank- note Money standard Lawful money Token money United States banking system Bank accounting Discount of commercial paper Credit money Easy money Treasury note Note issue Subsidiary coin U. S. Federal Reserve Board Call money Fiduciary money Federal Reserve system Representative money Rediscount Rediscount of com- mercial paper Free coinage Gratuitous coinage Limited coinage Letter of credit Cheap money Treasury loans Federal Land Bank Member bank	Mutual savings bank Bill of exchange U.S. regulation of interest rate Special interest account Private bank State bank State savings bank Clearing house Dear money Warehouse receipts Collections between banks Quantity theory of money Index number Mint par Mint charge Mint value Brassage Clearing house certificate Foreign exchange Rate of exchange Specie reserves Standardized dollar Stabilized dollar Bank for interest settlements Bimetallism Elastic currency Gresham's law Seignorage Limping standard Arbitrage Debasement of currency Par of exchange Secondary reserves Velocity of circulation Mobilization of credit Contraction of currency Inflation of currency Purchasing power par of ex- change Centralization of reserves Equation of exchange International flow of gold Gold exchange standard

DEGREES OF UNDERSTANDING OF BANKING

Average child	Bank as a place where money is safeguarded.
" youth	Bank as a place in which to keep surplus funds.
" householder	Bank as a convenience in paying bills - checking account
" thrifty person	Bank as a source of interest on savings or small loans.
" property owner	Bank as a source of large loans on property.
" investor	Bank as a source of loans on collateral.
" business man	Bank as an aid to business in short-time loans.
" college students	of money and banking.
" bank officials	represent various degrees of additional knowledge.
" expert bankers	of international banks.
" economists	specializing in money and banking.

ATTAINMENT OF PRECISION IN KNOWLEDGE OF FUNCTIONAL FACTS

more positive he is that he understands it fully. The expert who understands all the details listed above and others which might be added is probably much less certain that he understands banks and banking. The illustration affords a clear demonstration of degrees of understanding in the mastery of organized knowledge.

ATTAINMENT OF PRECISION IN KNOWLEDGE OF FUNCTIONAL FACTS

An analytical study of the trial results of the items in the Wesley-Gold, Kelty, and Pressey tests by Dr. J. R. Strayer revealed another peculiarity of the learning process. The following examples will illustrate this trait.

A budget is a

1. financial plan 2. bank statement 3. heavy club
4. cooperative society 5. monthly payment.

In the lower grades the pupils who had not attained the understanding of the term required by the correct answer tended to select "monthly payment" as the answer. In the later grades "bank statement" was the preferred error. By the tenth grade practically all of the pupils selected the correct answer.

The same point appeared even more sharply in cases where these independent workers had used the same test items, though with different options. Note the difference in the performance at the same grade levels. These percentages were obtained from the scores of several hundred pupils at each grade level.

Example 1.

Pressey

What is meant by legislation?

- (a) the passing of laws; (b) the convicting of criminals; (c) the collecting of money; (d) the enforcing of laws.

Kelty-Moore Legislation means

- (a) men connected with the government; (b) state body for the state only; (c) the making of laws; (d) the people who make laws; (e) a congress composed of two houses.

TESTS AND MEASUREMENTS

Scores:

Grade	4	5	6	7	8	9	10	12
Pressey	28	..	41	..	72	..	79	87
Kelty-Moore	15	24	15	18	26	23

The Kelty-Moore item is apparently too hard even for the junior high school, the score of the ninth grade being practically pure chance.

Example 2.

Pressey If fifty people vote unanimously for a certain officer, how many votes did he get?
(a) 25; (b) 49; (c) 0; (d) 50.

Kelty-Moore Unanimous means
(a) generous or unselfish; (b) the majority vote;
(c) having the agreement of all; (d) the people who vote the same way; (e) almost everybody.

Scores:

Grade	4	5	6	7	8	9	10	12
Pressey	75	..	82	..	85	..	90	90
Kelty-Moore	12	19	13	23	34	41

Here the Pressey item is so obvious that a pupil who had not noticed the term before might easily infer the correct answer from the question itself.

Example 3.

Pressey Who may be impeached?
(a) a business man; (b) a criminal; (c) a reporter; (d) a president.

Kelty-Moore Impeachment means
(a) a letter that goes through the mail; (b) bringing to trial a high officer of the government; (c) challenging to a duel; (d) a fraud perpetrated upon an officer of the federal government; (e) finding that an officer is guilty.

Scores:

Grade	4	5	6	7	8	9	10	12
Pressey	18	..	29	..	60	..	87	96
Kelty-Moore	13	15	10	18	46	38

ATTAINMENT OF PRECISION IN KNOWLEDGE OF FUNCTIONAL FACTS

In this case neither of the questions receives a score definitely above chance before the eighth grade. The Kelty-Moore question is, however, more searching.

Example 4.

Pressey Which is a mob?
(a) a fleet of battleships; (b) a decrease in prices; (c) an excited group of people; (d) a tax on imports.

Kelty-Moore A mob is
(a) many things bunched together; (b) one who feels superior to other people; (c) any large group of people; (d) people who crowd around a speaker; (e) a disorderly or lawless crowd.

Scores:

Grade	4	5	6	7	8	9	10	12
Pressey	66	..	75	..	92	..	92	97
Kelty-Moore	14	25	37	52	64	55

In this case the key word is not a "book" word but one which pupils have heard and seen often outside of school. Yet the low scores on the Kelty-Moore question indicate that the term is still imperfectly known at the end of the ninth grade.

Example 5.

Pressey Which adjective refers to the country in which one is born?
(a) alien; (b) foreign; (c) adopted; (d) native.

Kelty-Moore A native is
(a) Any person who lives in the country about which we are speaking; (b) a black person; (c) a person born in the country about which we are speaking; (d) a wild person, not civilized; (e) a cannibal.

Scores:

Grade	4	5	6	7	8	9	10	12
Pressey	43	..	65	..	81	90	96	..
Kelty-Moore	11	23	36	65	67	57

This is another instance of a word in general use outside as well as in the school. The difference in the scores, however, indicates that precision in its use is slow in developing.

TESTS AND MEASUREMENTS

Example 6.

Pressey

When is an injunction most often used?
(a) during a strike; (b) during a battle; (c) during an exploration; (d) during a voyage.

Wesley

An injunction is a
1. part of speech; 2. wreck; 3. union of two things; 4. court order; 5. form of advice.

Scores:

Grade	4	6	7	8	9	10	11	12
Pressey	7	19	..	42	..	55	..	69
Wesley	5	12	37	46	59	88

Neither of these questions demands a very accurate knowledge of the term. The difference in the scores is interesting.

Example 7.

Pressey

Which word means freedom from slavery?
(a) pardon; (b) agitation; (c) emancipation;
(d) exemption.

Wesley

Emancipate means to
1. enslave; 2. waste away; 3. liberalize; 4. make strong; 5. liberate.

Scores:

Grade	4	6	7	8	9	10	11	12
Pressey	22	25	..	59	..	69	..	86
Wesley	25	26	48	67	62	74

These two questions seem almost equally difficult. The added distinction between liberate and liberalize in the Wesley test accounts for the lower score on this latter.

In nearly every one of these examples, one version calls for a more precise understanding of the idea involved than does the other. In the first version only one or two of the options have any real connection with the question and the choice is more easily made. In the second version more of the options relate to the question, the choice is more difficult and therefore calls for more precise knowledge of the idea involved. In examples 2 and 4 the simple, incomplete, and superficial knowledge of the idea demanded by the first version is met by the majority of pupils as early as the fourth

WHERE PUPILS ACQUIRE FUNCTIONAL FACTS

grade. The greater precision demanded by the second version is not reached by the majority of the pupils before the seventh grade in example 4 and is not yet reached at the ninth grade in example 2. The difference is not quite so great in the other examples but is clearly marked. The following question affords an illustration of a degree of understanding beyond the reach of most twelfth grade pupils.

The people of a Congressional district are called the Congressman's (1) supporters; (2) opponents; (3) clients; (4) constituents; (5) henchmen.

These illustrations would suggest that the mastery of single ideas is, like that of highly organized ideas, a matter of progressive development not attained at any one grade level but built up in the course of a development extending from the fourth grade or earlier to the twelfth grade and later.¹

WHERE PUPILS ACQUIRE FUNCTIONAL FACTS

The cumulative learning of single functional facts revealed by these examples deserves further consideration. The terms used in the Wesley-Gold, Kelty, and Pressey² tests are virtually all in current use, appear in daily newspapers or current literature, are, in fact, part of the ordinary business of life. Or, to state this observation in another way, pupils who read current literature are exposed to most of these terms constantly. The selection of the items for the final test was accomplished by statistical procedure wholly unrelated to any specific subject matter offered in the schools. Items were retained or discarded on the basis both of progressive improvement in scores from the fourth grade to the twelfth and of the reliability coefficient of items within single grades. If the item showed steady improvement through all these grades, advancing sufficiently above mere chance, it became a preferred item. If it behaved erratically, it was discarded, even though the erratic behavior might be most significant of the effects of certain courses.

¹The phenomenon of progressive improvement in learning social material has also been noticed by Anderson and Lindquist of the University of Iowa. Cf. Anderson, H. R., and Lindquist, E. F.: *The Improvement of Objective Testing in History*, Second Yearbook of National Council for Social Studies. Pp. 100-104.

²See Chap. III.

TESTS AND MEASUREMENTS

The selection of items for the final form of the tests on the basis of continuous improvement without regard to the particular social science course offered at any grade level or, indeed, whether any course was offered, permitted the performance on the test to reflect social learning gained through other means as well as through courses in social science. This fact was fully confirmed by the results in the University of Minnesota High School, where the test was applied to all pupils in each of the six grades. It was confirmed also in other trials. There is nothing startling in this confirmation, however. It only serves to emphasize an important condition of social science instruction, namely, that such instruction is not the only avenue to knowledge of society.

The sources of social education are varied. They may be divided roughly into three classes, (1) personal experiential (2) formal class instruction, and (3) extra-experiential information. The first consists of the pupil's personal experience in his social life with his fellows and the more or less directed experience he receives at home, at church, and in various organizations and institutions to which he or his family are related. This type of material is most vivid. The pupil recognizes himself as part of it. It impinges upon his several senses and usually involves his active participation as well. These psychological advantages render this potentially the most important source of his social learning. Unfortunately the lessons from this learning are ordinarily expressed in local idioms, not in the universal language of literate society.

The third source is even more varied than the first. It consists of all those agencies for the transmission of remote, external information, the experience of others. It includes the pupil's general reading in books, periodicals, and daily newspapers; what he hears about the outside world, whether in conversation, from lectures, or over the radio; and what he sees as a detached observer either in passing, in pictures, or at the moving picture theater. Its relation to himself is not apparent; the experiences recorded are those of others, not his own. The impression made upon him is much less vivid, reinforced by fewer impacts upon his senses, involving fewer

WHERE PUPILS ACQUIRE FUNCTIONAL FACTS

of his feelings, and is usually fleeting, not persistent. The effect upon him will be largely dependent upon his own stores of apperceptive learning and interest. The third source springs from the ceaseless flow of apparently unrelated activities and ideas of the outside world. Its language is, however, of a more universal nature.

Formal instruction in the schools stands between these two. The material with which it deals is, like the latter, concerned with the thoughts and acts of the outside world. Unlike the latter, however, it is organized. The relationships have been worked out and the material is before the pupil in a permanent form to which instruction gives persistent force. The experiences with which formal instruction deals have many points of similarity and often relationships with the more ephemeral material found in the outside world. They also have many points in common with the more direct personal experience of the first source. The skillful teacher can gain greater vividness by pointing out these similarities and connections.

One of the chief functions of formal classroom instruction, therefore, is to serve as the link between the other two sources of instruction. The more efficiently this function is performed, the more effective will become the educative influences of the other two. Such instruction should enable the pupil to apply to the lessons of his own experience the universal terminology which the wider, grown-up world employs, thus enabling him to identify himself and his experiences with those of others in other parts of the world. It should also, by its careful tracing of relationship in wide human experience, enable him to recognize relationships in the ceaseless flow of otherwise apparently unrelated incidents of the contemporary world. It should bring him nearer to the goal of recognizing in innumerable personal experiences functional facts bearing upon the more remote operation of the larger social world; and to a similar ability in tracing the bearing of larger ideas of organized society upon the more minute affairs of his own daily life.

The relationship of learning in the schools to that from outside sources may, therefore, be likened to a chemical agent. It is the

TESTS AND MEASUREMENTS

means by which the other two are brought together. The two outside educative elements, the experiential and the extra-experiential, are operating always. They have begun before school days and they continue through life. Most of the material of extra-experiential learning is of the inert variety. The functional facts inherent in it are easily overlooked. This can be corrected in large measure by the formal school instruction which, according to its quality, more or less accelerates the instructional value of the other two elements as well. The evidence of these tests indicates that this acceleration occurs during school days and continues when the formal instruction has ceased.

RELATIVE PROPORTIONS OF INERT AND FUNCTIONAL FACTS IN THE SEVERAL SOCIAL STUDIES

After the first list of functional¹ facts, drawn from the text and reference books in social science up to the twelfth grade, had been compiled, an attempt was made to classify the terms. The task proved difficult because many terms belonged in at least two, some in three, categories. The more or less arbitrary classification showed approximately 45 per cent of the terms in political science, 33 per cent in economics, 20 per cent in sociology, with less than 2 per cent of ideas of time and space peculiar to history and geography. The more comprehensive list drawn from the junior college work in social science and from newspapers showed a slight shifting, sociological and economic terms increasing. A study of functional facts found in selected textbooks for grade 7-11 inclusive showed sociological terms 49 per cent, political science terms 42 per cent, economic 32 per cent, with the peculiar time and space ideas of history and geography together less than 1 per cent. The unusual total of percentage is due to the fact that over 20 per cent of the terms were classified in two, some in three, categories. These figures are of special significance in that, of the twelve textbooks analyzed, ten were textbooks in history, one in geography, and one in so-called unified social science. Nearly all

¹"Functional" facts as here used are limited to the restricted list as described pp. 8-13.

IMPLICATIONS OF STUDIES

the functional facts in history and a large percentage of those in geography are functional facts of social science. Those who analyzed the several types of material for the comprehensive list of functional facts found the books in history and geography high in "inert" facts and relatively low in "functional" facts, while the books in social science were very high in "functional" facts with a minimum—in some cases almost none—of "inert" material.

IMPLICATIONS OF STUDIES IN UNDERSTANDING

*The Probable Process of Learning Concepts, Ideas, or Functional Facts in the Social Sciences.*¹ The import of these observations would indicate that the learning process in the social science field could be illustrated somewhat in the following manner, the concept "state" as a governmental division of the United States serving to describe the general process.

1. The child has heard the name of his own state and of other states frequently before he ever goes to school.

2. Most children will have asked what a state is and may have received some form of dictionary definition, usually no longer than a sentence or two, which they grasp vaguely.

3. In the stories of the history of the country, incidents, events, and persons will be connected with certain states.

4. In geography, the distribution of resources, population, climatic conditions, occupations, and living conditions will usually be referred to in terms of states.

5. Any systematic history of the United States will cover the settlement of the states, the origin of their settlers, organization of their governments, the relation of the state to the national government, various incidents in each state's history, and the development of peculiar traditions and attitudes.

6. In the course in civics, the child will learn some details about the way in which his state government is organized and how it operates.

7. Meanwhile he will have noticed allusions to various states

¹See also pp. 14-16, 113-119.

TESTS AND MEASUREMENTS

in the newspapers, periodicals, and other news-disseminating agencies.

8. He will not have attained a working understanding of "state" as a governmental unit of the United States until he has learned wherein New York, and Nevada, Texas, and Rhode Island, Massachusetts, and Mississippi qualify as such states, while the District of Columbia and Alaska do not.

9. Most pupils will have attained this degree of understanding at the end of their high school course. If they continue in college past the sophomore year, some of them will have had the opportunity to take systematic courses in the history and government of their own particular states.

10. The average citizen charged with some responsibility for the conduct of affairs will continue to enlarge his knowledge and understanding of his own state and others, though not in a systematic way. If indifferent to politics, he will learn little, if anything, perhaps not enough to keep alive the functional facts that he learned in school. If professionally interested in politics, a politician, he will learn much about his own state and even others.

11. A few who are fascinated by the study of local politics or political science will devote their lives to a study of state government and will still have something further to learn about the subject when they cease working.

This illustration may be applied equally well to larger or smaller concepts. It can be applied to international finance, international politics, or even to so limited a problem as the relations between two individuals. In the case of the larger concepts, the working understanding will be attained later and by fewer individuals; in the case of less difficult concepts, earlier and perhaps by more individuals. In any case absolute, final, and definitive knowledge will never be attained, whether of international finance or the continuing relationship between two people.

Peculiar Nature of Ideas in Social Science. The illustration of the probable process of mastering an idea in social science suggests that there is a marked difference between the ideas taught commonly in the social sciences and those taught commonly in most

IMPLICATIONS OF STUDIES

sciences, technology, mathematics, logic, and philosophy. Ideas in social science are not definite in the sense in which ideas may be definite in some of the sciences or technology, for example. The physical sciences may present their ideas in precise formulae, and the formulae can be applied with equal precision, insofar as the scientist or technologist controls his material. In the chemical and physical laboratories, in the industries which fabricate raw materials, the raw materials can be refined or shaped in such a way that the formulae may be applied with almost absolute precision. The raw material of the social scientist is life itself over which he has very limited, if any, control. Most of his ideas are derived from the observation of society, and if he indulges in logical speculation for the purpose of explaining, improving, or controlling society, he still has the task of modifying his theory or formula to fit the actualities of social life. His theory or formula is valid only to the extent that it conforms to the ways of social life itself, or that social life can be modified.

Those ideas in social science which are derived from the observation of social relationships all represent a range of variation. They are derived from a series of recurrent, similar incidents, no two of which are exactly alike. Their statement in brief verbal formulae such as ordinary dictionaries employ serves to designate, but not to define. If the statement succeeds in characterizing the simple common denominator of the range of incidents involved, it may not cover a single actual occurrence of the idea. Equipped with a dictionary definition alone, the pupil might not be able to identify a single instance of the idea. He is like the finder of a ring of keys. To the finder they are only jingling metal like a dictionary definition. For the man who really owns them they will unlock many things. It is likewise with a social idea. A firm grasp of the idea is possible only if the pupil learns the typical range and the limits of the idea. He may begin his learning of the idea either by meeting it in his experience, whether direct or through reading, or by a verbal statement, a so-called definition. He does not possess an inclusive-exclusive knowledge of it, however, such as the scientist demands of a definition, until he has met the range of its

TESTS AND MEASUREMENTS

application, including the limits before which and beyond which it acquires a different meaning designated by a different term. The truth of this statement is demonstrated by the progressive improvement in the scores on the Wesley-Gold, Kelty, and Pressey tests.

Those ideas in social science which are derived from logical speculation about society in an effort to interpret, control, or improve society may be learned as verbal formulae to satisfy purely academic or intellectual demands. To become a part of social science, it is essential that they be derived from an accurate and comprehensive knowledge of society, and that they be effectively applicable to society. The task of the social scientist, then, becomes one of determining how much modification his logical formulae must undergo in order to be applicable to society. The illustration afforded by Montesquieu's speculation about the success of the English Government is pertinent. His theory that it lay in a rigid separation of powers, legislative, executive, and judicial, was expounded just in time to be taken over into the Constitution of the United States. The disparity between the fact that in actual practice they cannot be rigidly separated and the persistent effort to keep them separate has been a source of endless confusion not yet corrected after a hundred and fifty years. To meet the needs of social science understanding of such ideas would require a modification or qualification of the logical formulae to meet the needs of most, if not absolutely all, social relationships to which they apply.

Learning in the social sciences, therefore, is a constantly cumulative process which is only well begun by the time the pupil reaches the end of the public school system. Even the most simple of its ideas cannot be firmly grasped as a result of one incident or one statement of principle. Few of them are mastered even to the extent of a good working knowledge in any one grade. But their learning is begun as early as life itself and may be greatly accelerated by wisely directed instruction during the school years. This direction would seem to consist primarily of a judicious mixture of inert examples and functional facts, each teacher in the process

IMPLICATIONS OF STUDIES

aware of the examples already met, whether in previous instruction or in the business of life, and ready to widen the pupil's understanding either by fixing more firmly the functional facts already grasped or by initiating the understanding of additional functional facts.

Interrelationship of the Social Sciences. The social sciences have been separated into various departments for purposes of advanced study. The separation is necessarily artificial. No actual social event, no social organization or institution, can be classified accurately as exclusively economic or political or social. The larger or more complex it is, the more certain it is to have aspects of all three, and even more, categories. The difficulty which the investigators found in classifying even many of the simple functional facts in a single category is suggestive. A tariff is both economic and political and is certain to be social in its consequences as well. An industry, especially if large, is certain to combine all of these elements as well as many others. The upper officials will have most of their energies involved in problems of organization, formulation of policies and regulations, administrations, maintenance of morale, relations with other industries, with governments, and with the public at large, all of which are ordinarily classified as political functions. The lower officials and workers will have chiefly technical problems involved in the industry, though not without some political and some social concerns as well. The marketing of wares involves a consideration of social needs, desires, practices, as well as the more purely economic factors of transportation and storage. The same complex of factors is found in any organization or institution, whether its dominant function be social, economic, or political. It is equally true of state and church, industry and education, commerce and recreation. None of them can be understood fully in terms of one social science alone.

Though social learning is derived from many sources, most of them outside of school, the service of the social science subjects in the schools is one of economy in that learning. The economy in social learning accomplished in the schools is twofold. First, it may hasten the pupil's grasp of functional facts essential to his dealings

with the experiences of social life by supplying him with the typical range of illustrations involved in the functional facts. This should be of great help to him in dealing with the incidents of social life which he does not control. Second, it may widen the pupil's appreciation of the relationships in human affairs so that he will recognize more fully the extent of those affairs to which any given formula of interpretation, improvement, or control will apply and by which it will be modified.

The importance of the second of these services is not commonly realized. The universal desire to arrive at conclusions as quickly as possible interposes several important obstacles to such appreciations. One of these is the preference for sharply defined information. Whereas the social scientist, as the result of his long study, comes to the realization that specific events and specific institutions are merely devices by which society is working out its problems, that the underlying process is the important lesson, the beginner prefers the definite events and specific institutions. This obstacle is doubly exasperating because the beginner can only learn to appreciate the importance of process by learning a series of specific changes in society's solution of the problem, and because the social scientist has usually forgotten that it was in this way that he himself learned to appreciate the underlying process.

Another obstacle arises from the universal fascination with the possibility of a single explanation of human affairs. This urge for a philosopher's stone in the operation of society has in the past led to a variety of theories: a religious interpretation of history, illustrated by St. Augustine; the determining influence of physical environment as elaborated by Buckle; the economic interpretation of society in various forms from Malthus, Adam Smith, and Karl Marx to the many modern devotees; the political interpretation so eloquently expounded by Freeman and Seeley; and, most recently, the psycho-analytical interpretation derived from Freud by many contemporaries. It is only with the utmost reluctance that the social scientist himself has come to recognize that there is no one explanation of human affairs, that all of the interpretations which have been so enthusiastically advanced are but factors in the in-

IMPLICATIONS OF STUDIES

terpretation of large social phenomena, and that, in addition, there are probably many other factors which remain still to be identified. The beginner's weakness for the single explanation may therefore be condoned, but it is desirable from the point of view of the public welfare to wean him from that preference as soon as possible.

One goal of instruction in the social sciences is to extend the power of each individual to recognize the range of individuals and institutions which will be involved in any given program of reform or control, or even in any social activity of that individual himself. Nothing in past human experience justifies the hope that this goal will ever be fully attained. But human experience abundantly proves that society gains directly in proportion to its progress toward that goal, and no society has had a greater opportunity to approach that goal more closely than is afforded by the condition of practically universal high school education which now exists in this country.

One of the most important outcomes of learning in the social studies is the appreciation of the underlying process which particular institutions, specific measures, and definite events serve. It is also a fond hope of most social scientists that younger students may be led more rapidly, if not immediately, to the appreciation of the process instead of being satisfied with a knowledge of events, measures, and institutions. This goal, however, is only attained with very great difficulty.

This difficulty, or these difficulties, may be illustrated by the example of the railroads. For society the railroad has been probably the most important agency in the formation and maintenance of modern urbanized and industrialized civilization. The economist has recognized it as the chief device of the process of transportation in modern times. Most railroad workers, and many railroad officials, including even executives, boards of directors, and especially governmental regulating bodies, have been distracted by the immediate material elements of rails and cars, office buildings, and organization, to think of their work as railroading, pure and simple. Many of them may not have carried their thinking far enough in the first place to appreciate the importance of the process, while

TESTS AND MEASUREMENTS

others may have forgotten. The result of this limited view, when automobiles and, later, airplanes became efficient agencies of transportation, threatened to be very disastrous. Instead of welcoming these developments as aids in increasing the efficiency of transportation, railroad men frequently took a hostile attitude and resented them as interfering with railroading. It would be difficult to estimate the wastage in money alone caused by the struggle of railroad men to restrain the operation of auto trucks. Years passed before the more thoughtful railroad operators realized that their business was to aid in the process of transportation, and that this process could be greatly facilitated by linking both automobile and airplane to the railroad in such a way as to combine the peculiar advantages of each in the interest of society as a whole. Governmental regulating authorities have been slower still to recognize the situation. If this example illustrates the wastefulness of mistaking the institution for the process and the superior importance of realizing the underlying process, it equally well illustrates the difficulty of gaining and maintaining the idea of process.

Every field of social activity, however, affords innumerable illustrations of the difficulties that lie in the way of the attainment of the idea of process. The constable who, when electric lights were first installed in automobiles, insisted that drivers go out to the front of the car to turn on the lights furnishes another illustration. The old-fashioned judge who was so meticulous about the letter of the law and decided cases on that basis, rather than on the basis of the social process being served by the law, supplies another example. The school teacher who is intent upon the exact form of a new method and disregards the process of education which that method is intended to serve affords yet another illustration. The subordinate in any organization, economic, political, or social, who discharges his duties with a painful regard merely to the letter of his orders encounters the same difficulty in appreciating process.

These difficulties, so common in every field of social activity, suggest that the task of appreciating process is itself difficult. The attainment of the idea of process would seem to be achieved only through a thoughtful acquaintance with a vertical series of events,

HISTORICAL-MINDEDNESS

measures, and institutions. The material evidences of the events, measures, and institutions are grasped first—places, persons, and buildings. The immediate purposes which these serve are grasped next. The underlying continuous social process being served can only be grasped after a series of such material and immediate situations serving the same society at successive intervals are comprehended. The comprehension of process is therefore a feat of sustained thought, a complex abstraction, as much more difficult to reach as it is more abstract and complex than a simple functional fact. Every normal individual may be expected to attain to a limited comprehension of process in that field in which his thoughtful work lies. But the attainment of a wider comprehension of the numerous processes underlying the activities and institutions of a complex society can only be achieved by a very superior scholar in the field of the social sciences.

II. HISTORICAL-MINDEDNESS AND CULTURE PATTERNS

The list of values for which tests were projected included the term "historical-mindedness." This inclusion was indicated as desirable by the advisory committee of historians; and workers, both technical and subject-matter experts, were found to undertake the construction of a test. Historical-mindedness was defined as the habit of judging events and practices in light of the times in which they occurred. Thus, for example, the statement that a president of the United States kept slaves would be received in very different fashion in the twentieth century from its reception in the eighteenth century. Further analysis of the problem indicated that the attainment of this value definitely involved appreciation of the culture pattern of the time. Historical-mindedness differed from culture pattern as that term is used by the sociologist or cultural anthropologist only in the addition of the time element. The effort to test the student's possession of this value involved two steps, (1) knowledge of culture patterns, and (2) appreciation of the fact that culture patterns change in time. The second of these depended upon the first. Such a test, if constructed, would also have value for the sociologist who is concerned with differences in cul-

TESTS AND MEASUREMENTS

ture patterns according to place. The definition of the culture pattern would involve both time and place and the historian, too, would be concerned with differences according to place at any given point of time. The test of culture patterns promised to be of some service also to the economist and political scientist. Thus this value, which at first seemed peculiar to one field of social science, promised to be of general concern to all of them. Furthermore, the geographer's concept of "regional personality" is closely related to it.

Analysis of Patterns. The effort to analyze patterns of culture proved to be more arduous than was anticipated. It was found difficult to list the distinctive characteristics of a single period of history. Certain elements, such as inventions and discoveries, could be sharply differentiated as specific achievements. Their effect upon the life of the times, however, was a matter of gradual development and diffusion. It was not easy to determine how soon after the invention of gunpowder castles ceased to be effective fortresses and became merely residences, or how soon after the discovery of America this knowledge of new opportunities entered actively into the thinking of a large part of Europe. It was apparent that these effects would be felt differently in different portions of Europe. The historical pattern was thus affected by differences in region as well as time. Once this fact was recognized, it raised a corollary question whether distinctive patterns of one place at one time might not occur in practically the same or similar form at a very different time in another place. The manorial system of agriculture is generally regarded as characteristic of Europe in the Middle Ages, yet nearly every element which would be included in the pattern of such manorial life would appear in the eighteenth or even nineteenth century in eastern Prussia and perhaps elsewhere in Europe. Further study and reflection indicated that nearly every pattern of social life found in time might also be found today in some spot, the difference being largely in the extent to which that pattern applied, a pattern common to most of society at one time still existing only in isolated communities today. This would suggest that a test of historical-mindedness

HISTORICAL-MINDEDNESS

must be limited to some definite area and must include changes in the pattern from time to time in that area.

A consideration of patterns in limited areas revealed further difficulties. The same pattern did not hold for all people in the area. The pattern varied with the social classes, the upper classes being in touch with a much wider social world, while the lower classes lived in a very restricted pattern. The difference between these patterns was not sharply defined, for the classes fused into one another by almost imperceptible degrees until it seemed desirable to discover the pattern for each individual. Such degree of detail, however, was beyond the reach of any student within the years covered by this Investigation. No given area, even in American history, could have been studied intensively enough to afford the opportunity to test such knowledge.

Incidentally, the possibility of limiting the pattern to some phase of life or living was also canvassed. Institutional patterns such as forms of government, religious creeds, and economic systems were considered. Many of these had definite patterns, as, for instance, constitutions of states, or church creeds, which are set down in written form. Here, however, there was found to be a great difference between practice and theory. The pattern of a church creed and dogma as set forth in documents described the actual pattern of belief held by relatively few of the individuals who professed it. A given church might include within its membership some whose pattern of religious observance was almost fully described by the written dogma of that church. On the other hand, that pattern might definitely vary among the individual members of the church all the way from that point to the individual whose only point of contact with it was a nominal membership in the institution. The same range was found in the case of political governments and parties and economic systems. It seemed necessary, therefore, to conclude that the patterns described in such systems could not be used effectively in attempting to test historical-mindedness.

Anthropology has described culture patterns of primitive societies and recently geographers have developed the concept of

TESTS AND MEASUREMENTS

"regional personality" which is a pattern of physiographic and human activity elements. Both seem to offer an opportunity to determine such patterns with a completeness and precision required for purposes of test construction. The "regional personality" involving the relatively permanent geographical elements promised to be of real help toward the development of patterns required by historical-mindedness. Those regions which had been well populated over a long period of time and which had witnessed extended historical development, however, showed changes in the pattern of activities from time to time. It appeared, on inspection, that the more prominent in history a given region, the more likely it was to show such changes in activities and, consequently, in culture pattern. The less populous, less important in history, a given region, the more permanent the regional personality activity and culture pattern. This observation proved a further disappointment in the effort to test historical-mindedness.

Parker and Calkins, in their plans for a series of tests of geographical understanding, had included a test of regional personality. This proved, however, to be one of the most difficult of the tests in their series. Part of the difficulty lay in the definition of the complex factors involved in the several regional personalities. Perhaps even greater difficulty was found in formulating questions which would test this knowledge effectively. As a result, the test could not be completed in time to serve the needs of this Investigation. Such tests are contemplated in the program of the National Council for Geography and will doubtless appear in the near future. Meanwhile experience in testing the understanding of single ideas had involved so much effort, both in the formulation of the tests and in the statistical analysis of trial results, that the effort to test the more complex ideas involved in the culture patterns of historical-mindedness had to be abandoned.

Our limited experience might suggest that patterns, whether of action, thought, or circumstance, or the combination of all three in culture patterns, are chiefly theoretical. Whether laid down as programs or derived from observation, they do not exactly describe any individual or even group concerned in them. Though

SKILLS

this may be true of the complete pattern, it does not impair the value of the pattern as a guide to the understanding of its imperfect reflection by individual or group. It merely adds to the difficulty of using the pattern, calling for constant alertness in application and a readiness to accept incompleteness in detail as circumstances and conditions require. Awareness of such patterns as regional personalities, historical-mindedness, and other general culture patterns can be taught fairly early; proficiency in the application of these patterns can not be expected until late in the career of the student. He may select the most concrete and obvious elements in the pattern at first and gradually enlarge his use of additional elements as his knowledge increases.

III. SKILLS

Another class of values recognized as common to all the social science subjects was designated under the caption "skills." The appropriateness of this designation may be questioned, since the term is commonly associated with manual and mechanical processes. The tools and instruments employed in the social science subjects require mental rather than manual dexterity. Whether the applications of the term "skills" is a legitimate figure of speech in this connection is, therefore, doubtful. Perhaps "study activities" or "habits" may represent a more accurate choice of words, though neither is quite adequate. At any rate, this section deals with those methods of finding and treating information which are supposedly acquired in the course of these studies and are expected to be serviceable to the individual in dealing with similar information in later life. The acquisition of these methods of study may properly be likened to acquaintance with, and practice in, the use of tools in any more mechanical type of activity.

It may be questioned whether such skills should be considered apart from the materials with which they deal. Few pupils could be won to acquire skills as skills, and few teachers would feel justified in teaching them as such. Yet teachers realize that these skills are learned in connection with materials that are easily forgotten, and they expect the students to retain the skills for application to

TESTS AND MEASUREMENTS

new materials that are similar. A student is expected, for example, to know where to find information on technocracy and how to evaluate the information, even though the particular subject was never studied in class. Thus, from the instructional standpoint, skills have a large independent value and constitute one important class of outcomes of instruction in this field.

As in the case of understanding, the work of constructing tests to determine the pupil's mastery of these skills entailed much preliminary work. The first task was to identify the various skills required by the social science subjects. Secondly, it seemed desirable to analyze these into their component details which might be made the basis of test questions. Incidentally, there was ever present the question whether a given skill was peculiar to these subjects. A third task was to discover for which of these skills satisfactory tests were already available. The fourth step was to decide upon the most promising prospects for test construction.

The first task, that of designating the specific skills involved in the social science subjects was greatly facilitated by the work of Dr. H. E. Wilson, of Harvard University, who had long been interested in the problem of study activities. As a further aid to the Investigation, one of his students, R. A. Price, conducted a questionnaire inquiry of teachers throughout the country. Members of the staff of the Investigation and teachers of the Minnesota Seminar also devoted some attention to this problem.

The composite findings of these teachers and investigators may be generally classified in three broad divisions: (1) those concerned with the *acquisition* of information; (2) those concerned with the *evaluation* of information, and (3) those concerned with the *expression* of social learning.

The Acquisition of Information. From the simple acquisition of social information through sight and sound, the first type of skills moves on to include acquaintance with the devices and accustomed channels for the collection and distribution of social information. This involves such matters as tables of content and indices of books; cataloguing devices of libraries, such as card index and classification systems; standard reference works, encyclope-

SKILLS

dias, atlases, gazetteers, biographical dictionaries, or yet more specialized works, like the *Dictionary of American Biography* and the *Encyclopedia of the Social Sciences*; guides to periodical literature, including newspapers and more specialized references such as the Yearbooks, or the United States Census; bibliographical guides on fields of history, politics, economics, and sociology.

Ability to read might be considered as the fundamental skill of the social studies, but as this skill is understood by teachers in the field it is highly complex. The physical and psychological aspects of the skill are ordinarily not the concern of these teachers. Fortunately, in recent years psychologists have been devoting much profitable labor to a study of the entire reading process. The teacher of the social science subjects is, however, more concerned with its later stages. These involve knowledge of vocabulary, acquaintance with the symbols devised to summarize large bodies of information, and familiarity with the ideas that recur frequently. Ability to read efficiently is contingent upon such knowledge, and speed in reading, which at first glance seems merely a mechanical process, proves, upon analysis, to be a compound of apperceptive knowledge and ability to control eye-movement, the former being far the more important element. In his study of skills or study habits, Price found teachers listing "reading" as most important through the twelfth grade. It is the apperceptive knowledge essential to reading, rather than the mechanical process involved, which causes most of the difficulty in the later grades. The students find themselves halted by terms, expressions, symbols, and ideas which they do not understand. In addition, students have to learn to discriminate, as Henry Johnson pointed out in his *Teaching of History*, between different kinds of reading. The purpose and manner of the reading matter, as well as the purpose of the student in reading this material, are important conditioning factors in the skill of reading.

Ability to read is aided by yet another set of skills, namely, the practice of analysis and note-taking. The simple outline of paragraphs, the more detailed notes, the more closely organized outline of the larger reading units contained in entire chapters and books,

TESTS AND MEASUREMENTS

practice in filing notes from several sources dealing with the same topic, and practice in classifying and filing information from varied sources, all represent stages in the development of analytical skill. Teachers have devised aids in the development of analytical skill, such as outlines and syllabi, and guidance or study outlines. The ability to analyze, like the ability to read, is a complex skill dependent more upon the student's previous knowledge than upon mechanical arrangements. Like reading, its attainment seems to require the constant attention of the teacher throughout the secondary school years.

The reading and making of maps is a fundamental skill in geography, an essential skill in history, and more or less necessary to all the social sciences. It differs from the reading and analytical skills just discussed in that it obviously involves a technical vocabulary consisting largely of established conventional symbols. After these are mastered, the problem has many points of similarity to other kinds of reading and analysis. Knowledge of the meaning of the globe, lines of latitude and longitude, land and water configuration, and an understanding of the symbols for mountains, streams, lakes, and desert, cities, and states, and climatic features represent only parts of the special vocabulary demanded. In modern times the social scientist has imitated the geographer in devising graphic symbols and statistical charts for data of varying degrees of complexity. This practice of graphic representation of highly complex knowledge and massed data has led to much confusion. The use of pictorial devices commonly known has caused even children to assume that they understand the ideas involved because they recognize the pictures. The compression of large and complex bodies of knowledge into diagrams and charts which the eye can see at a glance has beguiled many with the illusion that they understand because they see. Such material, on the contrary, usually represents highly concentrated information and generalization with wide ramifications of detail which only those who know the detail can grasp. The principle of William James about generalizations applies here with special force. Ability to understand such charts and graphic symbols in social science is a matter of cumulative advance

SKILLS

in apperceptive knowledge, usually not achieved until the later years in college, and then in varying degree, according to the amount of knowledge at the student's command.

The acquisition of this vocabulary, so graphically expressed that the eye sees it at once, is also a slow process, far from complete at the conclusion of the secondary school. Apparently its mastery is also a matter of cumulative advance in apperceptive knowledge.

Skill in aural acquisition of information, though much demanded, has received relatively little attention from teachers or psychologists in recent years. The development of the radio and its increased use for the discussion of public affairs has increased the demand for this skill. It is one so constantly employed in instruction, almost exclusively in pre-school years, that its acquisition is probably assumed throughout. The wide condemnation of the lecture method by educational psychology some years ago has probably deterred teachers from consideration of this skill. As a result it receives attention only in those colleges where the instructors still insist upon examining the lecture notes of their students. It may be seriously questioned whether such neglect of this skill is justified.¹

The Evaluation of Information. The skill most difficult to acquire in the field of the social science subjects is that called critical ability. It is highly complex. It assumes proficiency in the other skills, has certain procedures of its own, and involves at every turn the application of judgment. It requires great refinement before it can be accepted as part of instruction, whether in school or in life. The process of this refinement consists chiefly in the application of rules of evidence as those have been developed in centuries of study. Those rules have been described for the treatment of oral testimony in law and written testimony in history. As such they serve all social science. The treatment of them by Langlois and Seignobos is generally regarded as standard. Each new device or method of disseminating social information develops supplementary refinements in the process of dealing critically with its material.

¹Cf. discussion of oral instruction in Horn: *Study of the Methods of Instruction*. Scribners, 1934.

TESTS AND MEASUREMENTS

Usually these are met by a modification of the rules already known, though some of these devices, such as the modern newspaper, the cinema, and the radio, have raised so many additional critical problems as to justify extended separate treatment.

Certain elements of criticism, however, are fundamental. These are of concern to both teacher and pupil. All social information is relative. The person reporting is quite as important as the information that he reports. Did he see everything? Was he qualified to appreciate or understand what he saw? Did he write as soon as he saw it or afterward—perhaps long afterward? Did he have any reason for reporting the event otherwise than he actually saw it, or for selecting only some particular aspect of it? In other words, the qualifications of the writer or reporter, the circumstances of the writing or reporting, and the possible bias of the reporter are elements which enter into the appraisal of the material reported, no matter what the medium through which the report is made. Next to the author and his circumstances, is the question of probabilities of correctness or error. Given the surroundings and actors in an event at a given time and place, is it likely that the account is correct? Does it conform in all details to the known details of the time and place? How much of the account may be accepted as certain, how much as conjectural? All of these are considerations which condition any account of social information.

The development of the critical skill is a long and arduous process and is expected to continue after the student leaves college. The skilful teacher can make important use of the practical learning which the student is acquiring in appraising information that he hears from his fellows. Even young children will have learned to take into account the expression on the speaker's face, his attitude, and the tone of his voice as well as some of the considerations already mentioned. The application of this same critical judgment to written and printed material is, of course, more difficult, but the same problems are involved. Comparison of new information with that already possessed is involved in both kinds of information, written and printed material, calling upon a yet greater body of learning. In both cases, knowledge of the person furnishing the

RANGE OF SKILLS

information, of his qualifications and his purpose, is necessary, though the written and printed material presents greater difficulties. Accurate knowledge of place, time, and circumstances of the scene and setting of new information, so essential to the judgment of probability, is also common to both classes of knowledge, though here again the written and printed material dealing with events more distant in time or space, or both, adds to the difficulty. The difference is largely a matter of degree and further refinement as the student's body of social information is enlarged, his ability to provide himself with the essential further information improved, and his alertness cultivated. It is a cumulative process taking on additional refinements of technique as the social learning widens and deepens from the horizon of the playground toward the world-wide interrelationships of economic, political, and social affairs.

Expression of Social Learning. While skill in obtaining information from oral sources has been neglected, even shunned, in recent years, there has been increasing attention given to activity of various kinds. This might be described as the development of skill in expression of social learning. The traditional forms of expression used in classrooms since there were classrooms, namely, oral recitation and written composition on social themes, have not received proportionately enlarged attention. In fact, the increased use of the new-type test has lessened somewhat the emphasis on those forms of expression. On the other hand, much attention has been devoted to the development of activities of great variety, practice in the use of parliamentary rules, practice in drawing, designing, reconstructing, collecting, dramatization, and re-enactment of scenes related to history and other social science subjects. All of these devices are intended to afford opportunity for individual self-expression and, incidentally, to cultivate added skill in that form of expression.

RANGE OF SKILLS

The following table affords some impression of the range of skills recognized in the teaching of these subjects, as it also indicates the judgment of teachers regarding the relative importance

TESTS AND MEASUREMENTS

and usefulness of these skills. Questionnaires were distributed to approximately seventeen hundred teachers and nearly five hundred replies were received. This table is taken from a preliminary report based upon the tabulation of the results on one hundred returns, selected at random, for each of the six grades indicated. Though confined within these grades, the tables throw some light on the practice of teachers both before and after these grades.

The following list presents the fifty-two items in the order of their appearance on the questionnaire.

A. Visual Activities.

1. Read to understand: Read to comprehend all of the ideas contained in the material.
2. Read to memorize: Intensive reading and re-reading of material to fix associations and allow immediate recall when desired.
3. Read to locate information: Relatively rapid skimming through a body of material for the purpose of locating and selecting specific data. Example: Look through the Constitution to find all the clauses that refer to the courts.
4. Read for enjoyment: Extensive rather than intensive reading which appeals to the pupils' emotions as well as to the intellect. Example: Reading historical novels and poems.
5. Study maps to understand all the ideas they contain.
6. Consult maps to locate specific information.
7. Study charts and diagrams to understand all the ideas they contain.
8. Consult charts and diagrams to locate specific items of information.
9. Study graphs and statistical tables to understand their full significance.
10. Consult graphs and statistical tables to locate specific items of information.
11. Observe pictures, scenes, models, relics, exhibits, bulletin boards, etc., for general impression and emotional enjoyment.

RANGE OF SKILLS

12. Observe pictures, scenes, models, relics, exhibits, bulletin boards, etc., to locate specific items of information.
13. Observe pictures, scenes, models, relics, exhibits, bulletin boards, etc., for specific details and emotional enjoyment.

B. Listening Activities.

14. Listen attentively to oral presentation of materials in classroom: Class discussion, declamation, or dramatization.
15. Listen to music in classroom: Listen to phonograph records or group singing.
16. Listen to radio programs: Listen to radio music, addresses, dramatizations, current events, etc. An activity impersonal in character, and which may involve a somewhat different psychological situation. It may be carried on at home or in school.
17. Listen to illustrated lectures: An activity involving both auditory and visual attention. There is some evidence that this combined activity requires a special study skill.
18. Listen to sound motion pictures: Both auditory and visual powers are involved and the combination may require a special study skill.

C. Oral Activities.

19. Read aloud. A combination of visual and oral activity.
20. Engage in group discussion: The activity is the reverse of listening to class discussion and involves participating in the discussion.
21. Give a special report or "floor talk." The pupil presents to the class a report of a topic which he has previously prepared.
22. Participate in a mock trial, a pantomime, or drama.
23. Engage in a debate: Formal argument and rebuttal between two or more students in connection with a controversial topic.
24. Give a memorized declamation.
25. Recite in class.

TESTS AND MEASUREMENTS

D. Writing Activities.

26. Write a descriptive theme: (A theme is here defined as "continuous written discourse.") It involves the collection, classification and adequate presentation of descriptive material in writing.
27. Write a narrative theme: An item of written discourse narrating a story or incident. It involves collection, classification, and adequate presentation of narrative materials in writing.
28. Write an expository theme: An item of written discourse explaining a trend, a point of view, a cause effect relationship, etc. Examples are book reviews and editorials. It involves collection, classification and adequate presentation of expository materials in writing.
29. Write poetry.
30. Write a play or pageant.
31. Write an advertisement or slogan or newspaper headline.
32. Make a summary or précis: A summarization or "boiling down" of material.
33. Make an outline or brief: A brief is an argument for or against a given point of view. An outline is a topical summarization of materials properly classified. Distinguishing between main headings and subheadings is necessary.
34. Write a question as for a "question box" or examination.
35. Write a letter.
36. Make a list of items: The pupil may be asked to prepare a list of the books he has read, the words in a lesson which are new to him, important personages, or the functions of the Supreme Court.

E. Drawing Activities.

37. Draw a true picture: Represent actual objects as accurately as possible. Example: Draw a picture of a mediaeval castle.

RANGE OF SKILLS

38. Draw a symbolic picture: Express events, objects, or ideas in symbolic form. Example: Draw a cartoon showing the results of the Civil War.
39. Draw a diagram or chart.
40. Draw a graph or statistical table: In addition to tabular presentation of data, statistical material may be presented in line graphs, bar graphs, circle graphs, and picture graphs.
41. Draw a map.
42. Draw a plan, design, or pattern.

F. General Activities.

43. Collect materials such as stamps, clippings, and old coins.
44. Prepare exhibits: A combination of many types of activity in a single project, usually on a cooperative basis. Materials collected by an individual or group may be displayed on a classroom bulletin board or in a special exhibit.
45. Construct models, stage properties, etc.
46. Take part in committee work: Participation in small-group work either in the classroom or in extra-curricular activities.
47. Take part in school or community projects: Such activities as participating in community celebrations, safety campaigns, school elections, undertaking work in school clubs, sharing in community celebrations.
48. Participate in the work of social science clubs, an extra-curricular activity.
49. Preside at a meeting of a class or club.
50. Take a field trip: An actual visit to a civic institution, historical locality, etc.
51. Make a survey: An inventory is made by the pupil as the result of investigating such problems as city planning, city housing, or school attendance.
52. Take a new-type examination.

RANK ORDER OF ITEMS AS ESSENTIAL

The column on the left contains order of list of skills. The columns which follow show the combined judgment of 100 teachers for each grade on the relative importance of these skills for the respective grade. The numbers in each case refer to the item in the first column.

Item	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12
1	4	3	1	1	1	1
2	21	23	19	22	18	17
3	5	6	6	7	3	6
4	13	13	17	11	14	22
5	19	23	24	21	15	24
6	3	1	6	2	4	7
7	25	29	29	29	22	19
8	6	7	7	9	7	5
9	27	26	26	34	18	17
10	10	9	16	16	10	8
11	17	17	14	17	20	27
12	16	13	11	13	12	19
13	19	19	18	20	25	35
14	1	2	2	4	4	2
15	49	41	49	47	38	47
16	49	45	37	34	41	35
17	33	36	43	27	38	38
18	44	49	45	44	46	38
19	25	26	26	26	28	38
20	1	3	4	3	2	2
21	14	11	10	8	9	8
22	36	34	37	39	50	47
23	44	41	39	34	33	30
24	49	45	49	50	50	51
25	7	5	5	5	6	4
26	29	32	23	19	22	27
27	33	36	32	23	29	32
28	18	23	21	15	17	10
29	49	51	49	50	50	51
30	49	51	49	38	50	51
31	49	49	47	50	50	42
32	8	15	9	12	10	13
33	10	12	8	6	8	11
34	29	34	34	33	35	35
35	36	45	44	50	44	47
36	22	32	24	18	31	25
37	40	49	45	38	46	43
38	40	32	41	44	37	47
39	27	18	21	25	20	21
40	23	23	31	33	27	15
41	10	10	14	9	13	13
42	40	38	37	44	24	43
43	44	38	41	38	43	38
44	33	29	33	38	35	32
45	49	45	52	47	45	47
46	15	41	14	24	29	15
47	18	16	20	28	24	23
48	40	29	34	43	31	31
49	33	21	38	31	31	29
50	25	20	29	29	43	25
51	40	41	39	43	43	41
52	10	8	11	14	15	12

RANK ORDER OF ITEMS AS TO FREQUENCY OF USE

Item	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12
1	10	7	5	1	1	2
2	38	34	29	35	25	19
3	6	7	2	11	5	7
4	2	2	9	9	10	13
5	42	40	40	33	27	39
6	4	2	7	5	4	3
7	44	42	43	40	38	34
8	10	9	8	16	4	9
9	46	41	44	42	37	33
10	10	15	20	18	10	6
11	25	22	12	14	23	21
12	17	15	14	11	18	14
13	13	12	17	14	17	21
14	1	2	2	6	6	9
15	48	49	51	48	44	48
16	35	37	40	33	52	26
17	29	32	38	33	33	36
18	51	51	52	50	50	45
19	27	25	25	24	30	28
20	2	5	6	3	8	4
21	4	2	1	2	2	1
22	32	30	26	42	40	42
23	33	27	19	22	22	14
24	50	47	49	50	49	49
25	7	5	4	4	5	5
26	36	34	24	24	25	29
27	37	30	34	26	35	40
28	42	37	34	20	16	6
29	45	45	50	48	50	51
30	38	39	45	44	47	49
31	40	49	47	46	51	46
32	20	13	11	13	15	17
33	17	18	9	9	12	12
34	14	22	22	29	19	29
35	24	34	36	38	40	43
36	14	27	17	17	25	24
37	33	43	39	27	47	43
38	22	36	29	27	28	36
39	27	13	20	18	14	16
40	29	29	36	37	21	17
41	10	10	13	9	10	19
42	48	47	29	44	44	47
43	22	18	29	29	31	36
44	20	21	29	22	31	32
45	41	46	41	38	47	51
46	17	18	17	22	20	23
47	17	25	32	41	39	31
48	16	44	48	47	33	36
49	27	20	22	31	38	26
50	31	25	33	35	35	26
51	52	51	46	52	44	41
52	10	11	14	6	12	12

TESTS AND MEASUREMENTS

EXISTING TESTS OF SKILLS

The effort to find tests of skills already prepared which might be used in the Investigation was vain. Exercises involving the use of skills were abundant. This form of test evidently antedated the new-type question form. Laboratory manuals in science, outline maps, and map exercises in history and geography contained numerous such exercises. Experience had, no doubt, indicated a considerable measure of success in their application. There were to be found, also, some efforts to apply the new-type test technique to the measurement of skills. Ability to read conventional symbols for massed data was being thus tested; and the Rochester Schools had devised a test for an even greater range of skills.¹ After extended consideration, it was decided to attempt the construction of tests which should be more economical than the laboratory exercises, and should probe deeper or more widely into the skills than the existing tests of the new type.

THE CONSTRUCTION OF TESTS OF SKILL

Realizing that this Investigation could probably not cover the whole range of skills developed or involved in the social science studies, it was decided to sample the two main divisions, the graphic and the more purely mental. The field of geography proved useful for the first and history for the second. Miss Edith Parker of the University of Chicago and R. D. Calkins of Mount Pleasant Teachers College, Michigan, collaborated on the first. Miss Marion Clark, supervisor of social studies, Montclair, New Jersey, and E. M. Hunt, Teachers College, Columbia University, undertook to devise tests of Critical Skills in American History. Miss Clark's experience is also recorded elsewhere in this volume.² Hunt found tests at the college level so cumbersome as to preclude the possibility of their extensive use for the purposes of the Investigation.

TRIAL OF THE TESTS OF SKILL

Miss Clark constructed a series of four tests for the elementary grades and four for the secondary school grades. Those for the

¹Cf. bibliography, under A. N. Gibbons.

²See p. 302 ff.

REFLECTIONS RESULTING FROM THE TRIAL OF TESTS OF SKILLS

elementary grades, upon statistical analysis, were found to satisfy the requirements laid down as minimal for objectivity, reliability, and grade to grade differentiating power, and, with slight revision, were approved by the Committee on Tests and ordered published. The critical problems for the secondary grades demanded more extensive revision, which has not yet been completed. Parker and Calkins completed the first two of the five or six tests which they had planned. These were submitted to trial and are published as purely experimental attempts. The tests were not designed to test any one program of instruction, but were aimed at the geographical knowledge and understanding that the average adult was expected to have.

REFLECTIONS RESULTING FROM THE TRIAL OF TESTS OF SKILLS

These tests have not been sufficiently tried to justify any positive conclusions. Their trial, however, did suggest a number of reflections that may be of service to later workers.

1. It is clear that a distinction must be made between those parts of a particular skill which represent what may be purely memoriter information and those portions of it which involve the practical application of such information to new situations and material.

The memoriter information, the meaning and relationships of conventional symbols representing massed data, the technical vocabulary of geography and social science, the rules of evidence employed in social science can all be learned as such. The possession of this knowledge can be tested by several forms of new-type questions with the same degree of efficiency as definitions or other memoriter material.

The practical application of such knowledge to new materials presents greater difficulties. The materials to which it is to be applied are so voluminous that the cost of the test is greatly increased, for the material to which the pupil is to apply his skill must be included in the test. Another difficulty presented in this phase of the test is that relatively little of these skills can be tested at the student's desk. Actual activity on the part of the student

TESTS AND MEASUREMENTS

is demanded which, of necessity, greatly extends the time involved in the test. The combination of larger amounts of material in the test and the longer time involved to allow for activities conspires to defeat the ends of efficiency essential to the new-type test technique.

2. Another reflection provoked by our experience may suggest to later workers a method of overcoming some of these difficulties. Supplementary studies including the submission of a simple completion type test to seniors in college suggest that the *memoriter material involved in these skills is not retained without actual practice*. In this particular instance, the college seniors were asked to complete sentences calling for the name of a particular reference device from which certain information could be most efficiently obtained. These students had made no special preparation for the test. On the other hand, the test called for nothing which an instructor might not expect them to know offhand at that stage in their careers. An individual check upon the findings of this test revealed the fact that, though the students had been frequently told where they might obtain such information, in high school as well as college, they had not been applying this knowledge in practice. In this case, the failure to answer correctly the questions based on memoriter knowledge also proved the lack of practice and, hence, the lack of skill.

Such experience raises the question whether a standardized test involving only memoriter knowledge could be made an efficient test of a particular skill. Would students, if they were led to expect such a test, not proceed to cram this memoriter knowledge and thus pass the test with only verbalistic information? This is done by most students with relative ease. Perhaps this possibility could be safeguarded by wording the question in such a way as to avoid the difficulty. After all, however, the real aim in a test of skill is to have the student demonstrate his ability *to do*. Furthermore, in doing, the student is expected to give evidence of the *habit* or practice of so *doing*. Only thus can the teacher be really satisfied that the student has gained one of the dynamic outcomes of instruction in this field. That, of course, can only be demon-

strated by individual performance under competent critical observation—a task beyond the reach of new-type test technique.

3. The highly prized ability to evaluate current information, the critical ability, is even more complex than the skill of finding information and the skill of employing graphic and conventional devices essential to the social sciences. This involves that universally esteemed, but ill-defined, quality of judgment. A person already possessed of exact knowledge on a given point is best qualified to judge of the accuracy of current information on that point. Such a situation, however, is seldom apt to occur. For most, if not actually all, people, the reception of new information requires evaluation on the basis of probabilities. Value is a relative matter, at least in social concerns. It involves comparison with other information bearing upon the same point. It is not unlike the practice of the skilful shopper who goes about from place to place comparing both quality and price. Applied to social information, judgment depends upon the possession of other information and other values to which the new may be compared.

Knowledge of the author's qualifications for the announcement of information may be determined if the author is known and information about him is available. If not, this appraisal requires more subtle calculation of probabilities, involving the publisher or publication through whom and by which the information is broadcast. It can only be conjectured in such a case, but there are degrees of conjecture.

Evaluation of information may be determined in large part by comparison with other available knowledge. Here again, the question of probability enters in, and this, too, is a relative performance -- the more accurate and detailed the knowledge, the more nearly correct the determination of probability. Evaluation of any information, indeed, requires comparison—antecedent knowledge of the places and persons with whom the information is concerned and a delicate determination of the probability of correctness contained in the current information.

In all cases, such evaluation, or judgment, presupposes the knowledge of a pattern of conduct and relationship constructed out of

past information. If the reader or hearer does not possess such knowledge, it is necessary for him to turn to the most reliable person or critic who does possess it. That pattern, again, is not an absolute one, but is constantly in a process of modification and reconstruction in light of additional authentic information.

All these qualifications tend to make the evaluation of new information a highly complex matter which neither a knowledge of established rules of evidence nor the possession of extensive antecedent knowledge can definitely determine. It is never a matter of absolutes, but always a relative matter. Under the circumstances, the use of the new-type test, presupposing as it does an absolutely correct answer, can have only a limited usefulness in testing this ability. *Individual performance under competent critical scrutiny would appear to be the best method of testing this value.* That is beyond the reach of the new-type test technique.

4. Skill in expression can only be demonstrated by actual example. Knowledge of rules of grammar does not prove ability to write effectively; ability to read a language does not prove ability to speak it; ability to interpret pictures does not demonstrate ability to draw or paint; knowledge of Robert's *Rules of Order* does not assure ability to conduct a parliamentary meeting; knowledge of economics does not prove ability to conduct a business successfully. All skills of expression involve muscular as well as mental knowledge, a knowledge that develops with training and practice. As these cannot be acquired without practice, they likewise cannot be tested without actual demonstration.

Perhaps more careful study of the stages in the development of such skills will reveal a definite progress similar to that indicated in the learning of ideas. If so, it may be possible to confine the test to a very limited performance upon some detail of more advanced stages in the skill. Successful performance of this detail might be a definite proof of all antecedent stages of the skill. The hope of maximum efficiency in the test of skills lies in this possibility.

5. In the test of skills, therefore, it would seem that the new-type test has only a very limited value. This judgment should not preclude further efforts to extend those limits. On the other hand, it

INTERESTS AND ATTITUDES

should warn against reposing too great faith at present in this method of determining the student's attainment of these dynamic values of instruction in the social science studies.

IV. INTERESTS AND ATTITUDES

Interests and attitudes are perhaps two distinct topics as viewed by the teacher.¹ They are here grouped together because, unlike information, understanding, and skills, both have an obvious emotional element not so apparent in the first three outcomes of instruction just considered. In the economy of school instruction, interest plays an important part as a means as well as an end. Attitudes and changes in attitude are of concern there chiefly as outcomes of instruction.

Interest. Recent pedagogical theory concerned with such methods as problems and projects and with the utilization of individual differences has given added importance to the element of interest. Research in vocational and scholastic guidance, and in economic and social rehabilitation has given further emphasis to this element. Teachers, whether aware of the activities of psychology and educational theory or not, have devoted much of their energy toward arousing the student's interest.

An exploratory investigation by Miss Lena C. Van Bibber into the specific difficulties encountered by classroom teachers of social science subjects affords some evidence as to the importance of interest in the minds of teachers. "Lack of interest" by the pupils was almost universally offered among the first three explanations for the difficulty of teaching certain specific topics or items. In citing this as a common cause of difficulty, the teachers were not only attesting an observed fact, but were also revealing their concern about the matter and implying efforts to discover ways of interesting their students in the topics.

Analysis of the element of interest for purposes of testing and investigation disclosed the great complexity of the problem. The psychologists tend at present to look for evidences of "sets" or ten-

¹Perry, R. B., *General Theory of Value: Its Meaning and Basic Principles Construed in Terms of Interest*. New York, 1926. Perry does not make this distinction, but includes both under the general term interest.

TESTS AND MEASUREMENTS

dencies, closely related to the biological constitution of the individual. The social scientist is more concerned with the external expression of individual interest as that reveals itself in attachment to particular social activities or concerns out of the great variety with which the individual is constantly surrounded. He may be described as primarily concerned with the possibility of modifying, diverting, or even changing the individual's interest. In this respect the classroom teacher shares the tendency of the social scientist and usually views as a deficiency in his teaching skill any failure to arouse interest in any topic in the day's work.

The discovery of any pronounced individual interest which could be utilized by the teacher, as well as the discovery of means by which interest could be aroused or diverted for the improvement of instruction were, therefore, both desirable objects of investigation. Pedagogical and psychological literature dealing with these questions were found to be abundant. It might almost be said that any discussion of methods of instruction dealt directly, or by implication, with some phase of the problem, as Dr. Horn's bibliography will show. For purposes of this Investigation, it was found to be impossible to set up a series of controlled experiments to probe a sufficient range of interests. For the sake of economy it was hoped that some form of new-type test might be devised which would yield a maximum of result with a minimum of effort. It was realized that this was a venturesome undertaking involving great effort without any definite assurance of positive results. Dr. Kelley was persuaded of the possibility of some success, and his conviction induced the Commission to sanction an effort to probe for some fundamental elements of interest as revealed in personality traits. With the assistance of M. R. Trabue, a test of such traits by means of a free association technique and extended statistical analysis was developed.¹ In addition, A. M. Jordan carried on an analytical study of certain traits by an adaptation of the methods employed by May and Hartshorne in their investigation of character traits. The report of Jordan's study is included in this volume.²

¹See p. 342 ff.

²See p. 437 ff.

INTERESTS AND ATTITUDES

Any attempt to find some simple, efficient test of the factor of interest can only be highly speculative at this time. The suggestions here offered are based upon a study of the literature and observation of the work carried on by the Commission, together with a few supplementary studies carried on by graduate students.

Individual variation in interest in any group of students must be accepted as an important conditioning factor in instruction. Just how far such individual interest is to be regarded as absolute and permanent is still a moot question. The pupil during his growing years undergoes important changes in social interest. The general trend of personality studies confirms this observation. Recent efforts to adapt the Strong Vocational Interests Test to sub-adult years tend to show that individual interests which are pronounced in adult years are markedly less pronounced in earlier years.¹ Whether this is to be interpreted as indicating that the individual has wider potentialities of interest at that earlier age, which circumstances of life tend to narrow and accentuate later on, or whether his further development would lead to such narrowing and accentuation of interest, irrespective of circumstances, can scarcely be settled now. While the observations offered in this paragraph deal with the psychological approach, the search for the fundamental organic predisposition of the individual, it must be noted that the evidence of such interest is sought in the individual's reaction to external stimuli. That in itself raises the serious question as to the relative importance of predisposition and external stimuli in determining results. Some psychologists would, no doubt, hold that such predispositions are definitely contingent upon, and modified by, external stimuli, and that, insofar as the theory is correct, such rigid separation of cause and result is not possible.

Social scientists and teachers of social science subjects would probably all agree that each individual has certain distinct personality traits. Usually this means a distinct combination of traits,

¹Fryer, D., *The Measurement of Interests in Relation to Human Adjustment*. New York, 1931. This work contains a comprehensive survey of efforts to measure interest and findings up to the time it was written. The particular reference regarding the adaptation of the Strong Vocational Interests Test to high school students is to the work of Dr. Rulon, now at Harvard University.

TESTS AND MEASUREMENTS

each of which occurs commonly enough in other individuals. The effect of such traits is noticeable in the individual's choice of activity. It is only in the most distinctive cases of individuality that those personality traits are expected to result in marked differences of behavior and activity. The question here involved is whether such differences in personality traits are sufficiently sharp in most individuals as to justify marked difference of treatment. It also raises the question whether most individuals are not capable of a fairly wide range in the expression or operation of interest.

Viewing the problem of interest from the standpoint of the materials of social science, which are concerned with the external stimuli, the situation is no less uncertain. There is, first, the difference between curiosity and sustained interest. Curiosity is natural about objects or situations, recognized as unusual or different from the accustomed. Curiosity may be fairly easily aroused about almost anything. But curiosity is only the gateway to interest. It may be easily satisfied and yield no residuum of interest.

Sustained interest is in itself a complex matter. To what extent is it dependent upon knowledge? Persons interested in games are not interested in all games. Is this difference due to the fact that they understand some and not others? And is understanding due to their interest, or interest due to understanding, or a combination of these elements? Teachers have applied duress of some kind to force indifferent pupils to study, and then have seen them develop a sustained interest in the subject. Students who achieve relative success in studies which otherwise have no interest for them have been known to continue that interest largely because of their relative superiority. Students have been known to avoid subjects in college because they developed a dislike for their prototypes in high school and frequently, when forced by college regulations to take some work in that field, have developed a great interest in it. The reverse has also occurred often enough to be remarked upon.

The business world presents examples of the operation of interest and even more of its cultivation. Sales resistance must be overcome. An industry engages in wholesale propaganda to achieve that end. Advertising is made as attractive as possible. Articles

INTERESTS AND ATTITUDES

ranging all the way from cigarettes to automobiles are pictured on bill-boards and elsewhere. They seldom are pictured alone. Pleasing colors are used, and the size of the picture is large enough to include an alluring background suggesting a hunting scene, or a vacation scene, or beautiful young women or men, an elderly man in an attitude of comfort, an elderly woman suggesting an idealized home—all elements arousing universal pleasant connotations. Or the advertisement appeals to a sense of social superiority within the price range of everyone. Evidently the merchant is convinced that he must make an emotional appeal to arouse the customer's interest, overcome his indifference.

Tom Sawyer's studied reluctance to allow others to white-wash his fence turned a punishing chore into a capital investment of considerable proportions. A librarian in a certain high school found herself deluged by demands for a manual of statistical information, never before the object of any general interest, because a young practice teacher had absent-mindedly held a copy of that book during the course of the whole recitation. The students were all eager to discover what rendered it so precious. These illustrations suggest that a sufficient degree of interest may be aroused in almost any enterprise, though devoid of interest in itself, to accomplish the expenditure of both effort and money.

These illustrations also suggest the importance of emotional factors in arousing interest. Whether the appeal is made to the student's pride of achievement, to his sentimental interests of one kind or another, his self-esteem, the prospect of fear or favor, punishment or reward, some initial appeal to the emotions seems necessary to overcome apathy or indifference. An instructor whose good opinion is valued, one who has a pleasing voice, manner, and personality, may gain a degree of interest in his subject of instruction from students hitherto indifferent or hostile to that subject.

It has been said that a teacher can impart any topic provided he make it concrete. Such abstract topics as the tariff and strict and loose construction of the national constitution have been used as illustrations. In both instances the exercise was worked out by having it involve something of fairly immediate concern to the

TESTS AND MEASUREMENTS

pupils or their relatives. The materials referred to were such as the student had experienced and about which he had definite feeling or about which he could readily generate feelings. Here again there was an emotional appeal. Furthermore, the term concrete is here used in a double sense. It not only refers to the fact that the illustration is "brought home" to the students, it also refers to the use of vivid, tangible objects such as the student might see, touch, and feel, or had so experienced. This raises the question whether the approach to an idea, an abstraction, or generalization, must be through the medium of specific, concrete examples which make appeal to several senses and thus afford a substantial base upon which the relatively unemotional abstraction may rest.

If the problem of interest is viewed as a factor in teaching, the great range of activities touched upon by the social science subjects should afford ample opportunity to enlist some degree of interest on the part of every student. History and human geography are as wide as the world in space and time and varieties of human activity. Both are rich in concrete material of time and place, persons and circumstances. Economics, politics, and sociology deal with special phases of life, but these phases touch the lives of all. It would seem possible for every teacher to arouse some spark of interest even in the most apathetic student, provided the teacher make the necessary effort to bring this material within the range of the student's comprehension.

Interest as the outcome of instruction in this group of subjects is expressed in a great variety of ways. The first evidence is the student's performance in the subject. Another expression of that interest is added study or reading beyond the bare requirements. Still further evidence of interest is afforded by the student's use of his leisure for further reading or study in this field. The choice of a career of scholarship in social science would be proof of great interest. A further and desirable expression of interest would be that adults should make use of social science studies to enrich their own activities, whatever these may be. And perhaps the highest evidence of interest would be that of general, studious, and active concern about public affairs.

INTERESTS AND ATTITUDES

Can such evidence of interest be tested? The performance in adult life is obviously beyond the reach of the classroom. A forecast of such interest can only be approximated. Most of this would be determined through scrutiny of the student's extra-curricular activities, voluntary membership in organizations concerned with public affairs, the use of his leisure for additional reading, study, or activity in this field. And of this only the alert and sympathetic teacher can obtain any extended knowledge. Within the classroom, the chief source of evidence of interest lies in the student's performances, whether in oral recitation, written essays, or essay type questions in examination. The range of interest and the application of that interest in the social science subjects is so great as to defy probing by a common test. Of the new-type questions, the completion form most readily affords the student the chance to indicate his study of anything beyond the common task. Such answers would be so varied even in a single class that only a competent teacher could appraise them accurately. Aside from this, the attainment of special interest in any field must be inferred from relatively higher scores on tests of knowledge in that particular subject.

At this stage in our progress toward the measurement of interest, whether as an element in instruction, or as an outcome of instruction, it would seem best to say that we do not as yet have a measure. The psychologist's probing into the fundamental personality traits has not advanced far enough to yield much that the teacher can use. The range of interest which the social science subjects can arouse or enrich is so great as to defy any attempt at a common or standardized measure. The importance of interest, both as a factor in instruction and as an outcome of instruction, has increased, and deserves added concern on the part of the teacher.

Attitudes. Attitudes, like interests, loom large among the expected outcomes of instruction in the social science studies and have received much attention from psychologists as well as from social scientists. It is generally assumed that social attitudes are the product of education and experience. Differences of taste and interest marking the individual are assumed to find their expres-

TESTS AND MEASUREMENTS

sion in any society. They may operate to incline the individual toward certain social attitudes within that society, but they constitute only one of a number of factors which determine that choice. Education and experience, the latter including the whole complex of environmental influences, are other determining factors.

There is a general belief that the social attitude toward the familiar is favorable, toward the unfamiliar unfavorable, whenever the familiar and unfamiliar come into conflict or even contact. This tendency is presumed to operate with increasing force as the individual grows older. Familiarity is, of course, primarily a matter of experience. It is also generally believed that education has the power to modify attitudes formed through experience and to generate or develop attitudes toward social practices, institutions, and ideas to which it introduces the individual.

The term attitude covers a wide range of meaning, only part of which falls within the field of social science instruction. Most of this teaching may be described as concerned with public affairs. Attitudes toward races, religions, nations, forms of government, economic systems, social classes and practices, as well as toward current social ideas and problems are among the more important tendencies usually recognized as falling within this field.

The term is often loosely used to denote preference, opinion, hopes, wishes, judgment, conviction, prejudice, and belief. Perhaps these terms express degrees of attitude. At any rate, the term applies to the mental or emotional set with which an individual approaches a social problem and which determines his line of conduct toward it. It is this dynamic quality which makes the determination of attitude so important to government, business, and society in general.

Psychologists have devoted much attention in recent years to the study of attitudes. Until relatively recent years their attention was devoted chiefly to adults, but in the past ten or fifteen years they have been applying instruments of attitude measurement to youth and children as well. Efforts have been directed chiefly toward studying the effects of various methods and materials on changes in attitude. The use of tests of attitude at successive grade

INTERESTS AND ATTITUDES

levels and in varied curricula promised to yield valuable results for the work of the Commission.

Examination of existing tests of attitude revealed not only the vast amount of attention which the matter had received from psychologists, but also disclosed several promising approaches which might be used by this Investigation. The attitude-scale technique devised by L. L. Thurstone promised to be most useful in the study of attitudes revealed at various grade levels. Other forms of attitude test were used in other parts of the Investigation concerned with adults. A series of issues was chosen by the Commission for the development of attitude scales. Thurstone arranged for a modification of the technique of this test to facilitate the testing of groups or classes of pupils. This procedure is illustrated by the following example:

STATEMENTS CONCERNING SOCIAL PROBLEMS

L. L. THURSTONE

FORM B

Name Boy or Girl Date

Age at last birthday School grade School

Following are a number of statements. Some of them will express your views and some will not. Place a check (✓) before every statement with which you fully agree.

1. On the whole, the prohibition laws are satisfactory.
2. It is better to suffer indignities than to declare war.
3. I am absolutely against any regulation of Sunday activity.
4. I do not believe in capital punishment but it is not practically advisable to abolish it.
5. Persons who offend against society deserve no pity.
6. The communist is too radical and extreme in his views.
7. The Chinese are superior to all other races.
8. Prohibition has been a great blessing to our country.
9. We should make every reasonable effort to avoid war.
10. Sunday Blue laws are ridiculous in an enlightened society.
11. Capital punishment is good because it keeps others from committing crime.
12. Only by extreme punishment can we cure the criminal.
13. We had better keep our eyes on Russia for a while longer before making up our minds about communism.
14. I hate the Chinese.
15. Prohibition should be retained at all costs.
16. War is hardly necessary in the modern world.
17. It is perfectly legitimate to spend Sunday for play.
18. Capital punishment is the only adequate punishment for murder.
19. We should abolish all of our penal institutions.

TESTS AND MEASUREMENTS

20. Communists should have the same rights as other people.
21. There is nothing about the Chinese that I like or admire.
22. There should be no legal restriction on the sale of liquor.
23. Peace and war are both essential to progress.
24. People who want to observe Sunday can do so without restricting other people.
25. Any person, man or woman, young or old, who commits murder, should pay with his own life.
26. Only humane treatment can cure criminals.
27. The communists are on the right road.
28. I don't like the Chinese.
29. Prohibition has corrupted our whole American government.
30. On the whole, wars do the world some good.
31. Even though Sunday observance is desirable, it should not be compulsory.
32. Capital punishment is absolutely never justified.
33. Harsh imprisonment merely embitters a criminal.
34. Communistic principles should be adopted everywhere.
35. I suppose the Chinese are all right, but I've never liked them.
36. Prohibition should not be forced upon a people whose majority oppose it.
37. Only a coward opposes war.
38. Whatever else a person does on Sunday, he should go to church.
39. No thinking individual can believe in capital punishment as a method of crime prevention.
40. Occasional offenders should be reformed.
41. Our government should be overthrown and a communistic one put in its place.
42. White people and Chinese can be good friends.
43. Beer should be allowed under government license.
44. The highest duty of man is to fight for the power and glory of his nation.
45. The purpose of Sunday is to glorify God.
46. Capital punishment is not an effective way to prevent crime.
47. A criminal should be first punished and then reformed.
48. Every communist should be put in prison.
49. The Chinese are a quiet, industrious, peace-loving people.
50. If the majority of states want prohibition, the minority should agree.
51. Every last one of us should refuse to take part in any way in any war.
52. Sunday Blue laws should be strictly enforced everywhere.
53. It doesn't make any difference to me whether we have capital punishment or not.
54. So-called scientific treatment of criminals does not stop crime.
55. Communism would destroy the family and the home.
56. The Chinese are a great people.

The application of these tests of attitude was surrounded with many difficulties from the outset. The issues involved, the statements of extreme attitude required by the scale technique, raised embarrassing problems which overtaxed the voluntary cooperation of even kindly disposed school administrators. It was, therefore, impossible to obtain as wide a trial for these tests as for the tests of other values. More serious than this, however, was the fact that most of the pupils did not understand the ideas involved and accordingly failed to register any attitude of significance. The issues

INTERESTS AND ATTITUDES

themselves were expressed in ideological form, and many of the terms used to indicate degrees of attitude proved unfamiliar and beyond the grasp of most junior high school pupils. The further evidence derived from tests of understanding indicated that it was doubtful whether the scales, as constructed, could be used effectively before the last year of the high school. Other of the difficulties involved in testing attitudes will be indicated in a consideration of some of the problems involved.

The value of an attitude from the standpoint of the social sciences lies in its index to social conduct, both passive and active. Is an attitude an opinion, a preference, a judgment, a conviction, a prejudice, or does it partake of all of these? If the last, attitude is not to be regarded as absolute but as a matter of degree. Thurstone's work indicates that attitudes vary in degree. Closely related to this question is that of the fixity of attitudes. Is an attitude once formed permanent, or is it subject to change? Does the relative permanence of attitude likewise vary, perhaps—to paraphrase a concept suggested by Bogardus—according to the emotional distance or intensity involved? Are attitudes which the individual associates with his sense of self-preservation more permanent than those associated only with his self-interest, and are these, in turn, more permanent than attitudes associated only with self-esteem and thus on to matters of no personal concern? The very fact that individuals vary in the degree of their attitudes might be taken to suggest a corresponding variation in the relative permanence of attitudes.

Psychologists who have worked with children in the pre-school years are convinced that many of the major social attitudes are established before the children enter school, certainly by the time that they enter upon the formal study of social science subjects. They have acquired a decided preference in the matter of church, political party, social system, some in race, nationality, form of government, economic system. Our experience in trying the Thurstone Attitude Scales with pupils in the intermediate and junior high school grades raises some interesting questions about the nature of such attitudes. How much does such an attitude involve?

TESTS AND MEASUREMENTS

Let us assume that a small boy has a pronounced attitude favorable to Republicans, opposed to Democrats. The source of the attitude is clearly the home. The father, perhaps both parents, are Republicans. The discussion at the home has frequently indicated the feelings of the parents on this point. The boy has come to associate adherence to the Republican party as bound up with his most precious loyalties. His attitude, accordingly, is one of strong partisanship for the Republican party.

The senior Robert La Follette evidently grew up in such a home and early acquired such an attitude. He never left the Republican party—even when he became leader of the Progressive group, it was the Progressive Republican group. Most of those who were to vote for him in his political career had grown up, like himself, with that passionate devotion to the Republican party. Yet Senator La Follette had relatively little difficulty in supporting many measures and principles opposed to those held by most Republicans in his day, some of them diametrically opposed. Nor did his followers have any great difficulty in supporting him in his stand. In this case, the attitude, though deep and lasting, seems to have been to the name merely. Similar illustrations could be found for every one of these early loyalties and attitudes resulting from them. The loyalty is to the name—not to what the name represents. The full pattern of the ideology is, of course, beyond the child's knowledge and comprehension.

The business world affords another illustration. A certain large hotel in one of the larger cities has changed its name twice during the past four years, though neither the building nor its management were appreciably modified. Preceding each of these changes in name the hotel had been the scene of a notorious crime widely heralded over the country. The notoriety was associated with the name of the hotel. Changing the name was thought to remove the notoriety, at least, for the casual traveler. Similarly, business houses, play houses most frequently, seem to feel that a certain stigma becomes attached to them if they retain the same name very long. Nearly every large city can exhibit several such houses whose names have been changed, frequently with the addition of New.

INTERESTS AND ATTITUDES

Or a newspaper or periodical identified in the public mind with a certain political or economic policy is acquired by a group of the opposite political or economic views, and the original name is jealously kept. Some business firms have made it a matter of business policy to avoid any unpleasant association with the name of their firm. They are at great pains to prevent the mention of their firm name in connection with any accident involving serious injury or any scandal involving one of their employees. The same principle is sometimes carried to a ludicrous extreme by zealous promoters who seek to establish some connection between their business and persons of favorable prominence in the public eye. Favorable or unfavorable sentimental associations with a name have a powerful effect upon people's attitudes toward the name. At any rate, business places a high material value on this belief.

One difficulty encountered in the effort to discover the pupil's attitude was an evident desire to avoid revealing a real attitude that he thought would not be approved. This reluctance is by no means limited to children. There seems to be a widespread feeling that attitudes are strictly personal matters to be divulged only when, and if, the individual chooses to make them known. Only the most rabid partisans are willing to divulge their attitudes freely, and for such no test is necessary. A large part of the population refuses to divulge its attitudes except under circumstances entirely acceptable to themselves. A stranger or an authoritative investigation conducting tests is not apt to be included among such acceptable circumstances. Here again, however, there is great variation. There appears to be no reluctance about revealing, even to a stranger, attitudes on matters of no great concern. But of what value is a show of such attitude which, because it is of no immediate concern, may be much changed if, or when, it does become a matter of great concern.

Another obstacle encountered in the effort to test attitudes was the operation of what the psychologists have called a defense mechanism, perhaps of compensation. Individuals somehow secretly guilty of socially disapproved conduct are at great pains to achieve a reputation of the opposite kind. Frequently they are most ex-

TESTS AND MEASUREMENTS

treme in public profession of an attitude which their own private conduct denies. Frederick II of Sicily, the Emperor, is credited with the enactment of the most extreme measures against heretics, though he was a man of practically no religious convictions himself. The essential value of an attitude as an index to conduct would be lost in a verbal or public expression of attitude in such cases. With persons of this psychological set a verbal test of attitude would yield results which must be interpreted as the opposite of their apparent meaning.

Still another difficulty observed in the administration of tests in some cases arose from the willingness of the subjects to oblige. It was apparent that some of the items in the test presented questions upon which the subjects had not spent any previous thought. Any answer, therefore, could represent only a flash judgment, the merest opinion, which further knowledge might seriously modify. Yet the subjects, pupils in school or adults, as the case might be, answered the questions in a way more indicative of a willingness to cooperate than of evidence of attitude. Answers of this kind are of relatively little value unless critically weighed.

These difficulties in the use of verbal tests of group attitudes, tend to make such testing procedure extremely doubtful in value, if not altogether useless, at present. Clerical scoring and statistical treatment alone would result in very erroneous impressions. To render them at all valuable, the papers would require scoring or weighing by some competent person intimately acquainted with the group tested. Thus, by far the most certain and valuable attitude tests have been those individually administered in psychological laboratories, a procedure too costly for purposes of general investigation.

Fortunately, the study of attitudes has been sufficiently pursued by many workers and from many points of view to yield some observations of value on the question of change of attitude. The psychologist with his individual technique, the historian checking the completed record of men's utterances and acts, the sociologist and political scientist in their studies of social control, the economist in his study of marketing, have found an abundance of mate-

INTERESTS AND ATTITUDES

rial, both on changes in attitude and reasons for those changes. Every large social institution, state, church, business, and almost every social activity of considerable dimensions has had to deal in a practical way with the problem of changing attitudes. Their experience, as well as the study of the social scientist, throws important light on the problem of changes in attitude and how it is accomplished.

Perhaps it may be assumed that a person develops some attitude toward everything that enters the range of his conscious experience. The intensity of the attitude, its consequent permanence, is dependent upon the emotional distance of the experience from the individual. The shorter this distance, the more nearly the sense of self-interest is aroused, the more intense and lasting the resultant attitude. This emotional distance is conditioned quite as much by the character of the individual as by the nature of the experience. Psychologists have observed the development of apparently persistent and fixed attitudes as the result of an experience that most individuals would regard as trivial. The occurrence of the experience at a moment when the individual was peculiarly sensitive was an important factor in the situation. The extent of the attitude, whether confined to a particular place, person, thing, or name, or whether extended to involve a whole complex of factors or association patterns, is likewise dependent upon individual variation, intelligence, and emotional set, as well as upon the concomitants of the experience.

As Walter Lippmann has pointed out, the fact that most attitudes have a verbal cue has been used extensively by agencies concerned with social control and business promotion. By little more than a change in terminology, legislators, administrators, and legal advisers have been able to make palatable policies and measures that once were regarded as full of unpleasant associations. Where the unpleasant associations are inextricably inherent in the policy or measure itself, these are bound to reappear, but meanwhile those concerned with the promotion of the measure will have gained their end temporarily, and that may have been their only object. Business has made use of this same situation more exten-

TESTS AND MEASUREMENTS

sively than any of the other social agencies, largely because its ends are so often immediate, such as the consummation of a sale. It is interesting to observe how frequently publishers or authors strive to approximate the title of some book widely and favorably known. The result is a flow of style in titles not unlike that in fashions of dress. It is clear that a change in name is one of the most widely used devices for procuring a temporary change in many attitudes.

Many objects of attitude, however, have names which are relatively permanent or are associated with cue words which cannot be easily changed. Those associated with places, or nationality, or race fall in this category. Those associated with established institutions, political, economic, or social, present the same difficulty. Individuals have frequently sought to avoid unpleasant attitudes by having their own names changed. They have succeeded only where it was their names alone that stimulated the unpleasant association and attitude. Towns occasionally and, in rare cases, countries have changed their names to avoid unpleasant past associations. For the most part, however, large and established organizations are dependent upon other devices than a change of name. The process of overcoming an unpleasant attitude is a correspondingly more difficult task. One of the devices much used in creating a desired attitude is that of building up a series of pleasant associations to counteract the unpleasant. Apparently it is easier to counteract pleasant than unpleasant associations, favorable than unfavorable attitudes.

The technique of war propaganda by which the emotions of a hitherto peaceful folk are aroused to a ferocious pitch makes use of a series of simple emotional appeals constantly reiterated. The enemy country is associated in the public mind with violation of the most sanctified institutions of family, church, and humanity. Stories of inhuman cruelty to prisoners, captives, especially women and children, of irreverence toward and profanation of religious shrines and dogma, of adherence to political and economic dogmas contrary to those esteemed at home have always been fairly effective. Not infrequently, extreme partisans on political and

INTERESTS AND ATTITUDES

economic questions seek to arouse a hostile attitude toward their opponents by accusing them of subversive designs upon the same sacred institutions.

The process of changing a deep-seated unfavorable attitude is much more difficult. Individuals, once the object of general unfavorable reputation, have sometimes succeeded in overcoming this by persistent publicity which presents them always in some benign, humanitarian, and human situation subject to the same sentimental concerns which afflict or affect the generality of mankind. The problem in the case of a hostile or unfavorable attitude toward some nation is more difficult still. This country had the interesting experience of being subjected to bombardment with propaganda by both sides in the early years of the World War. The difficulty in this problem is probably due both to the greater complexity of the situation and to the fact that there is usually a certain amount of counter-propaganda to sustain the unfavorable attitude. Oddly enough, a campaign of unalloyed laudatory and favorable publicity displaying the nation concerned always in a righteous pose is unconvincing. A gradual, apparently natural and undesigned portrayal of a foreign nation as afflicted with the same complex of good and evil, the same efforts to right wrongs and achieve justice, subject to the same or similar pleasures and sorrows as one's own country, human, in short, seems to be more effective.

The same procedure seems adapted to relieve unsocial race attitudes. *Abie's Irish Rose* served to take the edge off the tension reached in racial and religious intolerance, not because it sought to represent either race or religion as entirely righteous, but because it showed each to have both faults and virtues, to have the same fundamental sentimental concerns that affect humanity at large. Done with a laugh, it helped to end a situation which had become intolerable for most people. If it did not entirely end the unfavorable attitude, it helped to restrain it and helped also to permit a continued effort to relieve the unsocial aspects of the attitude. The task is a long one and a continuous one in dealing with

any large and established organization whose name cannot be changed.

The problem of changing attitudes is complicated by a natural tendency to generalize. If each conscious experience begets an attitude, economy of learning and knowledge begets a generalization. The individual is tempted to dispose of matters not within the line of his interests by classification. Thereafter he can avoid undue mental effort by tucking new experiences quickly into some established classification and thus dismiss them from his mind. The unfortunate aspect of this procedure is that, although the individual usually follows up an initial experience which is pleasant and thus learns more about it, he does not so follow up an unpleasant experience which has created an unfavorable attitude. He does generalize it, however, and make it a classification into which many other experiences having only a most casual similarity are quickly thrown to share the same unfavorable attitude. This natural tendency leads to much social injustice and harm. It can only be corrected through the interposition of outside influences which compel the individual to acquire added information about the experiences he would otherwise dismiss with his too ready classification and unfavorable attitude. The school is, of course, the chief agency of society to accomplish this result. Added knowledge, acquired as knowledge, should, in the natural course of time, result in an intelligent modification of the unfavorable attitude. Such, at least, is the hope and endeavor of educators.

Akin to the common tendency in adults to generalize is the tendency of children to form extreme judgments. An act or line of conduct is, to a child, either all wrong or all right; a person is either bad or good; if a person is once guilty of a wrong, he is certain always to do wrong; a person guilty of deceit on one occasion is thereafter to be regarded as always untrustworthy. Such judgments beget attitudes. These are not taken seriously because during childhood family, friends, church, and school are all laboring to correct them. Added knowledge is the best corrective. Thus the bright pupil, pursuing social science study through his school days, should have most of his judgments on matters of public concern

INTERESTS AND ATTITUDES

adequately modified. On the other hand, the poorer students and those who stop their work in social science subjects may continue to form such judgments or have their childish judgments crystalize into attitudes inimical to a happy society. The tendency to form such extreme judgments is corrected only by added learning. Their correction will be accomplished only in those fields in which individuals continue to work and learn. Since these fields are often remote from social science and public affairs, crystallized childish judgments may serve to guide them in their discharge of civic rights, obligations, and duties. Many important social institutions are thereby jeopardized.

The interplay of knowledge and emotion is apparent in all consideration of the change or modification of attitude. The emotional element is paramount in creating or fixing an attitude, which may take shape very quickly, even in a moment. The use of knowledge to correct or modify that attitude requires long and persistent effort. Such knowledge is most effective if it, too, has emotional elements. The more concrete and vivid that knowledge, the more emotion it will arouse and the more effective it will be in accomplishing the desired modification. Unfavorable attitudes, however, die hard; if the emotional intensity is great, doubly hard. Individuals possessed of unfavorable attitudes have been known, under the influence of some emotional experience, to profess a radical change in attitude only to revert later to their former attitude almost in its pristine strength. The repentent sinner at the revival meeting, the nationality hater after a stirring drama or motion picture, the persistent party man after a painful exposure of corruption by his party leaders, all afford excellent examples. A test of attitude administered to them at such a time might record a marked change of attitude. Such a test, however, would have little value in predicting future conduct. Modification of such attitudes can be accomplished only by repeated or reinforced treatment extending at least through school years and often beyond that time.¹

¹Perry, R. B., *op. cit.*, p. 526, suggests that there are four methods of modifying attitude or effecting "control of interests" which he summarizes as follows: "All of these methods are employed in everyday life—in the way of a man with a maid or of a maid with a man, in parental exhortation, in religious edification, in

TESTS AND MEASUREMENTS

The importance of the emotional factor in the formation of attitude is most fully witnessed in the efforts of church and state to establish permanently favorable attitudes capable of resisting all modifications. During the past century, the national states have displayed the most conscious and systematic effort to accomplish this.¹ Public ceremonials have been increased. The nation is symbolized not only by flags and statues, but often by some sentimental personification as well. Rituals have been established and more are contemplated. Symbols have been multiplied with greatly enhanced emotional appeal. National celebrations have been made occasions of public pleasure and delight. In some countries these are accomplished with pageantry in striking colors and varied uniforms of public officials. Stories of national heroes, thrilling narratives of national history, monuments commemorating heroes and national events, the history of the world from the standpoint of the particular nation are among the devices calculated to arouse an intensity of emotional fixation which no amount of logic and knowledge can easily overcome.

The development of interests and attitudes constitutes one of the major objectives in social science instruction. None of the objectives bear more definitely upon social conduct. There is every reason to believe that effective work is being done by the schools as well as by outside social agencies, both in the development of desired attitudes and in the modification of unsocial attitudes. The test of the effectiveness of special devices and methods in achieving such results, however, presents great difficulties. The desired goal of testing large groups by means of verbal test has not yet been attained. Under favorable conditions, the individual

efforts to extend personal influence, to obtain public support, or to create a public demand. In all of these cases the accepted maxims are the same. 'Give people what they want,' that is, present your case so that it will appeal to the object's existing interests; 'Show that you mean it,' that is, manifest in your own person the interest which you desire to awaken in another; 'People value only what they have to work for,' therefore, hold the object at a distance so that an effort is necessary in order to obtain it; 'See him just after dinner,' in other words, cultivate or seize upon a receptive mood when the subject is favorably inclined either through his general well-being, or through the absence of more attractive alternatives."

¹See Merriam, C. E.: *The Making of Citizens*. Chicago University Press, 1931.

COMPARISON OF NEW- AND OLD-TYPE TESTS

will indicate his attitude on almost any question, but such procedures are individual. This fact affords great encouragement to the psychologists who are engaged in devising general tests of attitude. The teacher of the social sciences, however, must probably be content to await much further work by the psychologists before an accurate general test of attitudes is developed.

V. COMPARISON OF NEW- AND OLD-TYPE TESTS

Nothing has been said thus far about the other forms of test not classified as "new-type" or "objective." The essay type of test question has been used, and probably is still chiefly used, by teachers in class work and, to a considerable extent, in examinations. It is used extensively, if not exclusively, in the various general examinations in European countries and forms a large element in the College Entrance Board examinations in this country. Beside the essay type question, these examinations also make extensive use of exercises and more or less intricate problems, even in tests of social science subjects.

These other types of question were not used in the Investigation for a number of reasons. They were lacking in the essentials of obvious efficiency. Only experts in subject matter could properly grade answers varying as widely in their content and method of approach as did such essays. Such experts, if available, would have involved greatly increased costs. Even with experts, the task of standardizing the values of such questions would have been practically impossible. Ordinarily, no two of the experts would give exactly the same, and only seldom similar, value to the answers, and the same expert would vary from time to time in his judgment of the value of an answer. Though such methods as are used by the College Entrance Board to insure uniformity in the grading of their papers doubtless serve practical needs, the demands of our Investigation could not be thus satisfied. Another deficiency of the essay question is the disproportion of time to the range of subject matter involved. The limited time of an examination permits only a very few essay questions. The student finds it necessary to write hastily throughout the examination period. He may be fatigued

TESTS AND MEASUREMENTS

by the writing more than by the thought, and he is so hurried that he has usually not time to form his sentences correctly, to say nothing of organizing his whole answer in well rounded form. Only a small fraction of the subject tested is ordinarily sampled by the answers. These three primarily mechanical considerations were sufficient to rule out the essay-type question and to minimize the use of the exercise- and problem-type examination for purposes of this Investigation.

The fact that the essay, exercise, and problem type of question were either eliminated or only slightly used as instruments of investigation does not imply that they are equally poor for purposes of examination, or for purposes of instruction. Our experience in the use of the more mechanically efficient new-type questions, limited though it has been, suggests that a comparison of these types of questions in terms of the ends of instruction in the social science subjects deserves at least a brief consideration. These ends may be treated under the categories already employed, namely, (1) exact information, (2) understanding, (3) skills, (4) interests, and (5) attitudes.

TESTING EXACT INFORMATION

Different groups of subjects vary greatly in the importance of precise information as an outcome of instruction. Demanding more precise information than literature, but less than language, the social sciences, when compared with natural science, logic, and mathematics, or technology, appear to require relatively little. Formulae in science, mathematics, and technology are learned in exact fashion in the classroom and are to be applied throughout life in substantially that same fashion. Ideas in social science, comparable to such formulae, can never be applied exactly in after life. Each application requires some modification, slight in some cases and more marked in others. The same quality of judgment which only the scholar in science or the master in technology is required to display in *some* applications of his formulae, every individual must use in *all* his applications of the ideas of social science. The reason for this lies in the fact that every social situa-

TESTING EXACT INFORMATION

tion is highly complex; that it contains unique elements or combinations of elements; that it is never the same as any previous situation—persons, place, or attendant circumstance usually, and time always, being different. The individual applying his social idea is always required to estimate how far the diverse and unique elements will necessitate modification of the idea. The scientist often, and the technologist nearly always, is able to control his material, so that his formulae may apply as learned, but the social scientist is practically never able to control the situations to which his ideas apply.

Whatever exact information the teacher expects the student to learn, whether this consists of specific persons, places, events, circumstances, definitions, conventional symbols, typical relationship, established procedures in gathering information or rules of evidence, is properly the subject of test at intervals during the year as well as at the end of the work. This information has the character of precision and definiteness. The answers are either right or wrong according to a standard or model furnished in the instruction. Both the new-type and essay form of question have been abundantly used in testing the student's possession of this knowledge. For this type of material, the essay question is cumbersome, wasteful of the student's time and the teacher's patience, and usually unfair to the student. If the teacher is forced to hunt through several pages of hastily written, almost illegible, scrawl for a few items of exact fact, it is doubtful whether the teacher will find them all, and, if the number of papers to be graded is large, it is certain that the teacher will not find them all. And what a waste of the teacher's energy in marking and of the pupil's energy in writing that small bundle of facts becomes! If proper and intelligent care is exercised in the construction of new-type questions, the teacher can test the student's knowledge of many times the number of items in the same period of time required by one essay answer. These answers will sample a much greater segment of the subject and therefore more correctly test this part of the student's attainment. And, in addition, this form of test will have involved much less fatigue for both teacher and pupil and greater accuracy in the

TESTS AND MEASUREMENTS

scoring. The comparison of the two types of test for the measurement of achievement of precise information is so decidedly in favor of the new-type test as to make the continued use of the essay form for this value seem a decidedly outmoded procedure.

Before accepting this result, however, a few *caveats* must be taken into account. The new-type test must be carefully constructed (1) to minimize the element of pure chance in the answer, and (2) to avoid ambiguities, thus insuring the student's unequivocal reaction to the true purpose of the question. The help of someone experienced in test technique is desirable. To insure the student's honest performance, the new-type test must be administered with even greater care than the essay type. Seating arrangements, varied forms for the same test, and efficient proctoring must be carefully considered. If circumstances require a modification of the test to be given at intervals, to insure an accurate comparison of the student's achievement, the several forms of the test must be weighted to insure reliability. This entails statistical treatment. Experience has shown the importance of these *caveats*. If they are not respected, if the teacher attempts to dash off such a test in a short time and administers it in a casual fashion, few of the advantages will be gained, and the incentive to dishonesty will be so increased as to outweigh even these few advantages. Such cautions cannot, therefore, be too greatly emphasized.

TESTING UNDERSTANDING

The understanding of ideas is generally considered to be a much more important outcome of social science instruction than any body of specific information. Ability to apply the idea correctly to a new or unaccustomed situation would usually imply understanding. Ability to select from a group of related ideas the one which most accurately meets a given situation would likewise indicate some understanding. Ability to trace the probable ramifications or relations of an observed social phenomenon would also afford evidence of understanding of the idea involved. Mere ability to repeat a definition of the idea would be unconvincing—especially if repeated in the exact words of text or instructor. It might be

TESTING SKILLS

then merely a verbalism having no more meaning than any other combination of words. The essay-type test has long been used, with varying degrees of reliability, to test understanding. The student's extended discussion usually offers sufficient latitude to reveal not only the understanding of the idea but also the degree of that understanding. If the new-type test could also be applied to this end, it would afford the teacher an opportunity to test a much greater range of ideas in a shorter time and thus afford a fairer and more complete indication of the student's attainment of understanding. The discussion of the Wesley, Kelty, and Pressey tests has indicated that this type of value can also be reached to a very appreciable extent by means of the new-type test. Discrimination between related ideas, ability to apply known facts to a new situation, and even some test of the student's ability to trace relationship, have all proved possible. The construction of such tests, however, is much more difficult, the possibilities of standardizing such tests less certain. It is not clear that they can be made to supersede the essay type of question for this value, but it is clear that they can be made a very helpful supplement.

TESTING SKILLS

The scholar's mastery of the skills of his calling is demonstrated in the successful execution of a thesis. This is comparable in the field of learning to the masterpiece of the craftsman in the manual arts. Most pupils, however, stop short of the higher degree in any of the social sciences. For the earlier stages such devices as a course or term paper, an individual topic, afford the student opportunity for the use of skill. All such devices might be described as exercises demanding the individual use of skills involving both knowledge and practice. They require extended time. Whether the new-type test can be developed to afford a more rapid, yet accurate, test of the student's possession of the skills remains to be determined. The two tests developed for this Investigation, the one on geography and the other on historical criticism, were essentially exercises which made use of new-type questions. Insofar as skill involves a certain body of precise knowledge, new-type tests can

TESTS AND MEASUREMENTS

be developed to test such knowledge. Insofar as skill is a matter of practice and use, it seems doubtful whether the new-type question can be of much help. The hint contained in one of the exercises mentioned under skill, that the knowledge essential to a skill is not retained except through practice, might suggest that the new-type test could be used fairly far in testing skill. A test of the knowledge applied sufficiently long after that knowledge was imparted might, if the knowledge were still present, give adequate evidence of the use of the skill. Much additional work, however, will be needed before the new-type test technique can be used except as a supplement to exercises.

There is one skill which may, indeed has been, seriously impaired by the excessive use of the new-type test. This is skill in expression, the ability to set forth some topic in social science clearly, convincingly, and agreeably. Teachers in college have begun to remark that students who come to them from school systems in which the new-type test has been used almost exclusively for a number of years are unable to express themselves cogently either orally or in writing. In school systems in which the new-type test has been used extremely, it has been possible for students to avoid writing a single complete sentence except in courses in English composition. Inasmuch as coherent and cogent composition is still one of the most widely used, as it is one of the most valuable, skills in social science, it would seem essential to continue to use the essay-type examination, if for no other reason than to afford practice in this skill.

TESTING INTERESTS

The discussion of interest as an outcome of instruction indicated that interest was strongly conditioned by the individual's emotional set. A test seeking to discover the evidence of interest in a given class would, therefore, have to provide for a wide range of interest. Such a condition would defeat the ends of efficiency which particularly recommend the new-type test. The essay-type question, which, by its very elasticity, permits the student freedom to reveal his thinking and reading, is probably the most economical

TESTING ATTITUDES

method of discovering evidence of a sustained interest developing out of the work in the subject. This type of question especially invites the creative and constructive interest of the good student. A proper appraisal and appreciation of this evidence requires the competent judgment of an expert—teacher, scholar, or both. It also involves the conscious employment of the essay-type examination as a test of interest.

TESTING ATTITUDES

The pitfalls of a direct effort to test the attitudes of students on questions which they regard as important and about issues of a controversial nature are so many as to defy any form of test. The new-type test is most direct and perhaps, therefore, least efficient in gaining results. The essay type affords the student an opportunity to reveal his attitude on many questions. The methods used by the historian in checking act against verbal statement may still be the surest method of determining real attitudes. Next to that, the essay type, as interpreted by the teacher, would promise to yield most satisfactory evidence.

THE EXAMINATION

The function of an examination given by a teacher at the end of a course to students with whom he has been working over an extended period of time may be purely ceremonial and is never more than a supplementary index of the student's attainment in the work. On the other hand, the function of an examination by strangers is to afford to the latter a faithful index to the student's total ability and attainment in the field. Though superficially similar, these two examinations are widely different in fact.

The latter, the examination as a true and faithful index to the whole achievement and ability of the examinee, is becoming increasingly important. The growth of civil service,¹ the development of personnel departments in industry, the use of qualifying examinations for admission to higher stages of learning, have all not only increased the number of such examinations but have given

¹Anderson, W., and Lambie, M. B., *University Training for the National Service*. University of Minnesota, 1931.

TESTS AND MEASUREMENTS

added importance to them. These developments have made it correspondingly necessary that the examinations used for this purpose be proper examinations.

The high hopes for a speedy, comprehensive, and accurate examination which were aroused by the application of the new-type test to measurement of learning have not yet been realized. This technique has proved efficient in testing precise information. It, therefore, has great value in examinations in technology, considerable value in the lower reaches of science and language, but as yet relatively little in social science subjects, primarily because the body of precise information in these subjects is of relatively less importance. Such experience as we have had in attempting to test the other values indicates that there is further possibility of improving the examination and of using the new-type test in that improvement, but there is much preliminary work still to be done. It is apparent that the improvement of the examination must receive the attention of social scientists as well as psychologists.¹

Meanwhile, until the necessary further work has been done, it seems desirable to caution those who are not expert against the extended use of the new-type questions for values other than that of specific information. Such tests are, of course, valuable in determining growth in social science vocabularies. The essay and exercise types of questions scored by competent teachers will be more serviceable until those further developments have occurred.

CONCLUSIONS

The study of testing in the social science studies will doubtless continue. The work of the Commission must necessarily end. Any conclusions offered at this time must be regarded as an inventory of progress at the point where the Commission ended its labors. No finality is or can be claimed for them. Further work will certainly modify them and may prove some of them wrong. They are presented here as our best present judgment.

¹It is interesting to note the advance which has been made where such cooperation has been systematically applied, e. g., the test in United States history in the fifth annual Iowa academic contest, 1933.

CONCLUSIONS

1. *Are Tests Necessary?* In the past generation it has been seriously questioned whether the test or examination is a necessary or desirable adjunct to instruction. The main arguments of the discussion for and against the examination are briefly summarized in a recent work¹ and need not be reconsidered here. The following observations may be added. It is not likely that pupils will cease to desire to submit the evidence of their learning at intervals, both to satisfy themselves on the certainty of their progress and to gain the friendly, helpful criticism of those who are masters of the learning. It seems no less likely that conscientious teachers will continue to reassure themselves by test of the satisfactory completion of one stage of the work before entering upon the next stage. Some form of periodic testing will probably continue to be regarded as an essential element in instruction within the classroom. Outside the classroom, the demand for testing has greatly increased in recent years. The growing centralization of school systems has accentuated the importance of the test as an inventory of educational efficiency, not only in the minds of the school administrators, but also of the public. The tendency of colleges to limit their enrollments, whether as a matter of total enrollment or for admission to more advanced or specialized work, has given added emphasis to the test as a selective device. This development has led some college administrators to seek the perfection of tests which can be used for guidance purposes to help the student determine his own course with the least amount of waste.² Perhaps even more important has been the increase in civil service and the development of more searching selection of employees for industry. Testing has thus been extended to the world of affairs, giving added emphasis to the testing done in the schools. These observations add heavily to the conclusion of the writer cited that testing has become more, rather than less, important. It is not a question whether there shall be

¹Jones, E. S., *Comprehensive Examinations in American Colleges*, New York, Macmillan, 1933. Cf. Kelley, T. L., *The Scientific Method in Education*, Revised Edition, Macmillan, 1932.

²Paper read by Dean J. B. Johnston of the University of Minnesota at a meeting of the American Council on Education and the Educational Records Bureau, New York City, November 3, 1932.

TESTS AND MEASUREMENTS

tests, but rather how these tests may be more effectively adapted to discovering whether purposes and ends—objectives—have been realized.

2. *Are Existing Tests Adequate?* The growth in the use of testing has been accompanied by a scrutiny of tests used. The customary examinations derived from classroom practice were assailed as unreliable, often invalid, and highly inefficient.¹ They were criticized as inadequate to the objectives of instruction, and their grading was denounced as highly inaccurate and subjective, scores frequently being determined by irrelevant personal preferences of those grading the examinations. This criticism was so well founded and mounted to such proportions that some institutions of learning began to make even their general examinations a matter of sustained study.² The new technique developed by psychology, with its obvious advantages of efficiency and accurate scoring, commended itself rapidly to the industrial world and appealed strongly also to educational administrators. It seemed so superior to the traditional examination procedure that it threatened to replace the older method entirely, and psychologists were invited to develop tests for nearly all school subjects. Standardized tests employing this technique and fortified by statistical studies of reliability and validity came into general use. These, too, have now been used long enough to disclose weaknesses almost as serious as those found in the older procedure. The conclusion that existing tests are not adequate is one of the most definite conclusions derived from our experience.³

3. *The Relation of Tests to Objectives.* Analysis of objectives, whether of all of the social science studies *in toto* or of any one of

¹The literature criticising the traditional examination is now quite voluminous. See appended bibliography.

²Harvard and Swarthmore are pioneers in this field. The University of Chicago is engaged on a very extensive program at the present time.

³Perhaps the best brief summary of the deficiency of existing tests from a technical standpoint is the paper prepared for the American Council on Education by E. F. Lindquist of the University of Iowa. This paper was based upon extended study and is supported by a number of graduate theses, among which is an unpublished doctoral dissertation by E. C. Denny, *An Investigation of the Defects and Weaknesses in Certain Objective Tests in American History*. University of Iowa, 1932. This has special interest for this field.

CONCLUSIONS

them, disclosed certain definite types of objectives, each type involving important divisions as the objective was considered in connection with specific material. The types of objectives may be indicated as follows:

- A. A body of information
- B. A body of ideas accompanied with special vocabulary
- C. Ability to apply ideas to new situations
- D. A body of skills
- E. The development and acquisition of interests
- F. The acquisition of certain social attitudes
- G. Ability to express social learning clearly and effectively

While each of these types has many subdivisions which vary according to the particular subjects taught, the types occur in all of the subjects. Any examination purporting to test learning in this field must take at least these types of objectives into account, however much it may vary in detail. Few, if any, of the existing examinations do so. The new-type examinations with their strict limitation to the material set forth on the test papers may be more deficient in this respect than the old procedure which, by its very elasticity, permitted the examinee to reveal much more learning than was called for by the questions. Both, however, are subject to serious criticism for failure to consider the range of objectives.¹

4. *The Interrelation of the Types of Objectives in Social Science Subjects.* The objectives in social science subjects have a definite relation to each other. All are related to information. Ideas are generalizations dealing with specific facts, whether as descriptive of known and studied facts or as applied to new facts. Skills are skills in dealing with information. Attitudes, and here we are referring always to *cultivated* attitudes, are attitudes toward certain facts or bodies of information. Social information constitutes the focal nexus of all objectives. While this is true, the order of relationship is the opposite of the obvious. The possession of information does not assure the possession of the other objectives. The possession of the cultivated attitudes would more nearly imply the

¹Osburn, W. J., *Are We Making Good in the Study of History*. Public School Publishing Company, Bloomington, Illinois, 1926.

TESTS AND MEASUREMENTS

possession of the information. So, too, would the possession of the requisite skills. The aim of efficiency in testing, namely the goal of testing as much as possible in as little time as possible, is not served by a test of information alone. The number of items of information which can be tested in any given time is relatively small. These items can easily be committed to memory without acquiring any of the other objectives. The process of acquiring the other objectives is a laborious one involving much information and reflection about the relation of that information. The process of gaining the information is essential to the acquisition of the skills, but skills themselves are the result of much practice in this process. Testing for the objectives of instruction in the social science subjects must take this relationship into account.

5. *Suitability of Different Types of Question to Test Types of Objectives.* The different types of objectives vary greatly in definiteness. Information is most definite, attitude least so.

A. The possession of information¹ can be best tested by means of the new-type test technique. This type of question can search out a greater range of information in shorter time and be marked with greater accuracy at less cost than can the essay type. Any of the forms of the new-type question may be used for this purpose, multiple-choice, matching, completion, rearrangement, even the true-false form. Those questions having the greatest definiteness and the least opportunity for chance are to be preferred. The completion form is least definite; the true-false form offers the greatest opportunity for pure chance performance and is, therefore, least desirable.

¹Those who have followed the discussion closely up to this point may have been somewhat troubled by apparent changes in the distinction between information and other forms of knowledge which have appeared at various stages in the progress of the work. In a sense, these changes have faithfully reflected the growth in appreciation of the problem involved. The distinctions are not universal, but arbitrary. As arbitrary distinctions they have a practical value in the teaching of the social science subjects, which is the only justification for their employment. The differences involved upon which the practical distinctions have been made are matters of degree rather than of kind.

As here used in the concluding portion of this discussion, *information* refers to the materials of the social world, persons, places, and objects of various kinds and the terms by which they are identified or described. The external circum-

CONCLUSIONS

B. The possession of ideas, so far as this involves merely vocabulary and fixed definitions, can also be best tested by means of the new-type test forms. The same considerations apply to the selection of forms as in the case of the test of information.

C. The ability to apply the ideas to new situations can be only partially met by the new-type forms. The ideas vary in complexity and the situations calling for their application involve varying amounts of description, some beyond the scope of the short-answer form. Of the new-type forms which may be used, those calling for a high degree of selection are to be preferred. The four or five-option best-choice form of questions, in which each of the options bears some relation to the question, proved most efficient for our purposes.

stances or events may also be included. Information, so considered, is descriptive, referring to separate and static materials.

Idea here refers to the relationship of the materials of the social world, actual and potential, their purposes and uses, the problems and processes of their relationship. Any institutions, principles, laws, and ideals of society are, of course, included in this category.

A distinction of practical value to the teacher of social science subjects may be applied to both *information* and *idea* as here defined. All of the *information*, and so much of the *idea* as consists merely of verbalisms designating or describing ideas, institutions, verbal formulae, laws, is only potentially valuable to instruction and learning in this field. Such knowledge constitutes the means but not the end or purpose of instruction. If the learning is not carried beyond that stage it is, so far as the teacher of the social science subjects is concerned, "inert."

This distinction is also arbitrary and perhaps of practical concern chiefly to the teacher of these subjects. Such material is an essential preliminary to the "functional" knowledge with which the teacher is most concerned. Perhaps in an absolute sense such knowledge is not "inert," the difference being again a matter of degree. Most teachers of these subjects, however, would regard their efforts as unavailing, if their pupils possessed only a list of names and verbal formulae without realizing the interrelationship of the materials and ideas which these names and verbal formulae designate. The distinction, therefore, must be regarded as practical rather than philosophic.

It is probably true that no item of information which the pupil acquires is completely "inert." It is equally probable that no one ever attains complete functional knowledge of social information. The process of endowing what the teacher calls "inert" knowledge with its functional attributes is a gradual and continuous one. It proceeds more rapidly in the case of information in which the pupil is vitally interested and less rapidly where the interest is little or non-existent. The perfunctory acquisition of information in the latter case may consist largely of "inert" material. In any case, the point in the process of developing information from the almost inert to a distinctly functional knowledge is not easy to establish. It may be less a point than a span, before which the information has seemed inert, and after which it is recognized as functional. It is this functional addition to information which the teacher recognizes as understanding.

TESTS AND MEASUREMENTS

D. The possession of the essential skills calls for performance in addition to knowledge. Where the performance can be accomplished by means of an exercise of small compass, the new-type forms of question may be used to supplement the exercise. In this case even the true-false form may be used efficiently, where the intermediate operation of the exercise is assured. The more advanced stages of skill, however, require a scope of performance beyond the technique of the new-type forms and demand the application of expert judgment to insure the best appraisal of the performance.

E. The evidence of developed interests, which are matters of both degree and kind and are therefore less tangible than any of the preceding objectives, is less easily discovered by means of the new-type forms of question. Of the new-type forms only the completion form affords a sufficient range of answer to sample both kind and degree of interest, and this requires expert appraisal. The essay form of question affords still greater elasticity and it, too, must be appraised by expert judgment.

F. The possession of the cultivated attitudes is least tangible and most elusive, and the test of their true presence is afforded perhaps more often by indirect or incidental evidence than by direct. The ordinary forms of the new-type test are too direct to be serviceable. Whether psychological research will be able to devise some efficient instrument to test this objective remains still to be determined. The essay form with appraisal by expert judgment seems at the present the most certain form of test for this objective.

G. The ability to express social learning and understanding, clearly and effectively, is a matter of performance orally or graphically. The written essay or exercise appraised by expert judgment offers the best test of this objective. The use of the short-form type of test for this value is definitely objectionable.

6. *Modification of Tests to School Grade Learning.* Those concerned with the testing of learning in the social science field are fully developed adults accustomed to the use of that learning on an adult plane. Though psychology has relatively little definite

CONCLUSIONS

guidance to offer, there is reason to believe that social learning follows some order of development and growth. The obvious biological development of the child suggests important changes in attitudes and interests. In this period he is transformed from a condition of complete social dependence to one of apparently complete social independence. From one who seeks guidance, he becomes a person who tends to rely entirely upon himself. These facts strongly condition social learning. Our experience in applying the same tests to successive grades in school suggests certain phases of progress in social learning which later research, it is hoped, will specify more closely.

A. The accumulation of information is a progressive process. Allowance should be made for this fact. On the other hand, where the information is organized, as most materials of instruction are, the main facts are most thoroughly learned and progress is marked by the addition of subordinate information. If the time for testing the pupil's store of information is limited, the test should consist primarily of subordinate items in various degrees of detail. If the pupils know the latter, the knowledge of main items may be assumed, while the reverse is not true. A test made up chiefly of main items is, therefore, not a test of the accumulation of information.

B. Ideas are learned progressively. This is just as true of a single idea as of a store of ideas. Only the very simple ideas are mastered with precision in the earlier grades. Imperfection is to be expected and the questions used in the test, if of the new-type, should provide for degrees of precision.

C. Ability to trace relationship is the product of teaching as well as of study. Awareness of, and ability to trace, multiple relationship is a slow development and not to be expected except as the result of definite training. In the case of highly organized activities, such as those of a great government or industry, pupils may be led relatively early to recognize the fact that anything affecting the center of the organization will sooner or later affect all parts of the organization. In the case of unorganized activities, such as that of a new invention, the recognition of the full sweep

TESTS AND MEASUREMENTS

of its relationship requires both knowledge, imagination, and training. The public official who was able to profess adherence to the slogans that the interdependence of the world should be recognized, that national industries should be protected, and that Smithville money should remain in Smithville affords an interesting illustration of a lack of such training in recognizing multiple relationship. Limited results are to be expected from pupils still in school, but these can be tested by essay questions.

Organized institutions, governmental, economic, or social, all represent multiple relationship, the larger and more highly organized, the more numerous and extended their relationships. The aim in instruction is to have the pupils appreciate this multiple relationship. The step by which the pupil advances to that learning is through the concrete and visible appurtenances of the institution. The material equipment, buildings, the personnel, relation of personnel to material objects, relation of personnel to other persons, first in single relationships, and after that in widening relationships, appear to enter the field of the pupil's appreciative knowledge in about this order. Knowledge of the material characteristics of an institution can not be used to imply knowledge of the institution. Knowledge of some of the relationships of the institution implies much more, but is still inadequate. Knowledge of the institution is not fully satisfactory until the pupil can trace the multiple relationships of the institution in at least its more important economic, political, and social phases. Such knowledge can only be approximated during the public school years. The multiple relationships of unorganized and sporadic elements in society are more difficult to trace, since imagination and thought as well as learning are involved. The thought concepts behind institutions, the pattern of ideas involved in the dogma or ideology, represented usually by words ending in -ism, are more difficult still, because they lack the concrete material and the activities of the institutions themselves. The notion of process, which regards institutions and ideologies as passing devices through which society is seeking to satisfy its fundamental needs, is most difficult. During school years, therefore, the tests should be designed to

CONCLUSIONS

measure the degree of such knowledge of relationship, appreciation of each *additional* relationship marking desirable progress.

D. The acquisition of skills is not separate but incidental to the growth of knowledge. The simple prototypes recognized and cultivated in the earlier grades become increasingly complex as the pupil builds up the store of his knowledge, both of facts and their relationship. The additional refinements, though recognized under separate names, owe their value to the soundness of the foundation of antecedent work in skills. Tests of skills, therefore, require provision for degree of mastery. This is usually afforded most easily by offering degrees of difficulty in the material upon which the pupil's skill is tested.

E. Interests are conditioned by a number of factors including individual inclination, biological development, personal experience, and learning. Assuming that the chief duty of instruction in social science subjects is to cultivate interest in public affairs—those social concerns which affect the group as well as the individual—the conditioning factors exercise an important effect. Lacking personal responsibility, the pupil's interest is correspondingly difficult to develop. His economic dependence, his physical and social immaturity, make it difficult to advance far in the cultivation of a well-rounded interest in all phases of public affairs. There is enough similarity between his personal group-life experiences and the political affairs of more adult society to develop an interest in this phase of public affairs somewhat earlier. The development of interest in the other phases must necessarily lag, even though teachers seek to stimulate that interest by vivid, concrete experiential material. Evidence of interest detected in incidental revelations in the pupil's work should be fostered.

F. Attitudes and judgments might appear to undergo an opposite development. Positive and decisive at first, they are gradually modified under the influence of increased knowledge. As in the case of interest, one of the first concerns of the social sciences is to develop attitudes and judgments on public affairs. The attainment of attitudes and judgments based upon well-rounded knowledge of society is the goal of that instruction. Teachers may assume

TESTS AND MEASUREMENTS

that their primary function is to help develop that knowledge, and with more or less assurance that, if that task is well done, the attitudes and judgments will take care of themselves. Under those conditions, the possession of the socially desirable attitudes is not to be expected until toward the close of formal instruction. The evidence of extreme judgments and attitudes in the earlier years is to be expected, their improvement a matter of gradual development until the pupil has acquired a wider or deeper knowledge of society as a whole.

7. *The Possibility of Developing Standardized Tests.* The obvious interests of efficiency would suggest the desirability of standardized tests equally useful in any part of the country year after year. Statisticians have devised a method for such standardization. However, even the construction of a test of factual information, to be valid as well as reliable, is an expensive undertaking much beyond the resources of the individual teacher or subject-matter department. That fact alone would make it desirable, once a test had been made, to continue its use widely and for an indefinite time. It remains to consider the factors involved.

A. Specific information lends itself most easily to the new-type form of test, and hence to necessary statistical treatment. The amount of specific information which may be used in the social science studies, however, is so great that only a small portion of it can be used in any single course. Localities will inevitably draw upon their own resources for some of this information. The pupil learning about society in London will do so through a different set of facts from those learned by the pupil in New York. The pupil in Boston will learn many facts which the pupil in California will not learn, and vice versa. If it were possible to construct a test limited to those specific facts which all four pupils learned, it would consist of few items and those only the general items. Thus it would cease to be a test of progress or a test of the extent of information. If a test of specific information which did measure extent of information were devised, it would have a maximum efficiency in one geographical area, and that efficiency would decrease more or less with the distance from that area. As for time,

THE PROBABLE LEARNING PROCESS IN SOCIAL SCIENCE

even in the area of maximum efficiency such a test would continue to be so efficient only if the same body of facts continued to be used, and if there were enough reliable forms of the test to preclude the possibility of memorizing and cramming for the test.

B. Ideas are more nearly universal. A great variety of specific information can be used to teach the same ideas. The pupil in London, New York, Boston, and San Francisco would acquire a fairly large body of common ideas. Thus far, however, only the single, less complex ideas have lent themselves readily to the new-type form and the corollary statistical treatment. This limits the extent of the usefulness of such a test as a measure of advance in learning. Again, it has been found that there are many ways of expressing the same idea and that different ways are preferred in different areas. There would be consequent variation in maximum efficiency in different areas. The same area, moreover, changes its preference in terms for the idea from time to time, and this would render the test less efficient in any area at later times. It might be argued that widened social intercourse renders the standardization of the vocabulary of social ideas increasingly important and that, whatever the local preference for variety of expression, this should be accompanied by a knowledge of the standard term. This form of test, therefore, offers the greatest possibility of standardization.

C. The other values in social science instruction have not yet lent themselves sufficiently to the new-type technique to consider the possibility of standardized tests. Psychologists and social scientists may ultimately devise such tests, but at this writing too much work remains to be done to afford more than a hope.

VI. THE PROBABLE LEARNING PROCESS IN SOCIAL SCIENCE

This discussion would be incomplete if it did not consider the learning process in the social science field, even if little can be set down with assurance at this time.¹ Psychology has concerned itself but little with what it calls the higher thought processes

¹See also pp. 14-16, 45-53.

TESTS AND MEASUREMENTS

and still less with the higher thought processes in the social sciences. Educational measurement, though it has done some pioneer work in this field, has as yet advanced little beyond the pioneer stage. It is doubtful whether substantial progress will be made on this problem until the social scientist and expert in educational measurement combine in genuine cooperation. Meanwhile the following suggestions may be offered as the result of the cooperation this Investigation made possible. These suggestions are of varying value. Some can be stated almost with the force of definite conclusions, while others can be advanced only as conjectures which some years of teaching experience, observation, and attentive study have justified. As such it is hoped that they may stimulate others to more definitive research. They are presented with due reservations.

1. The social sciences are expected by society to have practical applications. They are derived from a consideration of social phenomena and are intended to be of service in the conduct of human affairs. They are functional subjects, that is, they are meant for use to mankind. They are functional in two senses. They are functional in a passive sense when their service is to satisfy general curiosity, give added pleasure in helping to explain matters which come to the attention, and allay needless fears about the possibilities of new developments in a dynamic and changing society. Whatever they add to the cohesion of society by supplying common ideals and traditions may be included as part of their passive function. They are functional in an active sense when their service is to inform, modify, or direct conduct. No tests of conduct were attempted in this Investigation, though several of the tests implied effects on conduct. For the social scientist the thought that a knowledge of social science is a conditioning factor in social conduct represents its chief functional value, one of the goals of his efforts in teaching.

2. As a matter of conduct, the child begins to learn about society almost at birth and continues to learn more or less throughout life. Such learning is truly functional, acquired through activity and expressed through activity. It is the learning derived from

THE PROBABLE LEARNING PROCESS IN SOCIAL SCIENCE

experience with society and is concerned with the direct contact of the individual with his social environment.

3. The rapidity with which the child acquires this knowledge and retains it is, in part, dependent upon his ability to identify his experiences and to extract from them those elements which are apt to be of use to him in the future. The identification of such elements by distinctive names is the most economical device which society has discovered for this purpose.

4. The child is aided in this process by the older members of his social environment who have had similar experiences and who supply names for the essential and functional elements of the experience. They may or may not always be at pains to supply him with the names most widely used for these elements of experience. These more widely used or standard terms the school is expected to supply.

5. Much of the child's early school days is spent in learning the terms which society has found most generally satisfactory to express the functional elements of the experiences which he has already had. Pronunciation, spelling, reading, and writing are all involved in this process. The acquisition of the names lags far behind the experiences themselves in his early school years and is still some distance behind at the end of his schooldays. Perhaps it would be more correct to say that the gap between experiences and the appropriate terms by which to identify them is never fully bridged.

6. Such experience is, in his earlier days, limited almost entirely to personal contact and sensory perception. It is incomplete in connection with those materials, persons, and acts which come within his range propelled by forces beyond the reach of his senses. These more remote elements, though part of his own experience, he must learn from others who have been able to trace similar experiences back to their sources, or apparent sources. The service of the social science subjects in this connection is obvious.

7. The accumulation of the functional elements of his own experience constitutes the core of his knowledge bearing upon conduct. What he learns about the more remote factors affecting

TESTS AND MEASUREMENTS

matters which require his action must become connected with this core to become functional, that is, to modify his conduct. Social science instruction will succeed in its aim of conditioning conduct only to the degree to which it succeeds in establishing this connection, attaching the more formal learning of the classroom to this active core.

8. There appears in general to be some regular order in which the child acquires functional knowledge from experience. He recognizes, first, the concrete elements—artifacts, persons, and action—and identifies them. Their relationship to himself, their relationship to one another, their relationship to more remote concrete objects are appreciated successively, and the identification of this relationship follows the same order. The rapidity with which the individual child acquires this knowledge depends upon his ability, his interest, and his instruction.

9. His learning about matters more remote, beyond the reach of contact, follows a similar order, except that it is preceded by words, instead of being followed by words. As a rule he acquires first the verbalisms, then the material connotations of the words, then the functional ideas. His appreciation of the functional ideas depends upon the extent to which he realizes the concrete experience from which it is derived and this, in turn, upon the closeness with which that concrete experience can be related to his own similar experiences.

10. The acquisition of functional facts from the experience of others is an extended process. Left to himself in reading or hearing about the experiences of others, the pupil will gain the concrete elements in varying amount according to ability, interest, and vividness of presentation. The functional elements of that experience will escape him unless especially emphasized. Repetition and thought are necessary.

11. Precision in the grasp of a functional fact requires a number of considered illustrations of its occurrence, preferably at intervals. Under the most favorable conditions, several illustrations are required, and usually many more are needed, to give the pupil an understanding of any term as it is commonly used. Should he

need this term for active conduct, even more would be required. The result of such tests as those of Wesley, Kelty, and Parker and Calkins indicated that, even in the case of relatively simple functional facts, some pupils did not attain until the twelfth grade the degree of precision that other pupils had acquired as early as the fifth or sixth grades.

12. Pupils of lower ability can be taught functional facts derived from remote experiences, but they require more illustrations, greater emphasis, and a longer interval of time than do the brighter pupils.

13. The effort to teach functional facts by definition and thus to hasten the process of imparting the world's experience is illusory. Until their meaning is fortified by experiential material they are mere verbalisms, and, as verbalisms, are soon forgotten. The wider and deeper the experience of the pupil, the more receptive he will be to definitions and principles.

This conclusion suggested by the trial of several of our tests is corroborated by the findings of the Modern Foreign Language Investigation, as also by the Classical Investigation. Their problem was concerned with the relation of the rules of grammar to the functional use of the language. Both found it a mistake to teach formal grammar before the pupil had had experience with the language. The emphasis of the rule in connection with actual instances of its occurrence proved more effective.

14. Organized relationships such as are involved in schemes of government, organized industry, social institutions, and highly developed activities are compounds of functional facts. At first exposure, the pupil will grasp only the main nexus of this pattern, which will have for him only the value of a simple fact. The whole complex will not be appreciated until late in his school days, if then. Thus a bank is merely a bank, the detailed ramifications and significance of which for the processes of economic activity are only learned later.

15. A system of ideas, like a pattern of relationships, is grasped first as a single idea and usually requires some material association to be retained, frequently by imaginative personification. Symbols may be required as aids for the beginner.

TESTS AND MEASUREMENTS

16. A person's standard of knowledge about any body of organized relationship is the amount of knowledge of it which he possesses. If he has any knowledge of it at all, he will assume that he has complete knowledge, until his knowledge of it is too great to be recalled in a moment.

17. A person forms opinions upon which he is ready to act easily in response to the dictates of apparent self-interest and momentary expediency. To base such opinions upon knowledge and to qualify his action by knowledge is an ideal of whose attainment the individual is most certain in inverse ratio to his actual knowledge. The more knowledge he has, the less ready he is to act on impulse.

18. A person usually generalizes his own experience and observation, and is not ready to modify that generalization except as additional modifying information is forced upon his attention.

19. A person, left to his own devices, will confine his learning to those matters which fall within the range of what he regards as his own interest. All other matters will be quickly disposed of under some convenient classification as matters of indifference or aversion.

20. Matters of learning not within the field of direct personal interest will be regarded as definite, final, and permanent, and the possibility of change in such matters of learning will be ignored and even resisted.

21. Interrelationship is appreciated with difficulty. Experiences are treated first as separate entities. Their relation to other experiences are next appreciated as single relationships. An organization, as, for instance, an army, is first grasped as a series of single relationships, from that of private soldier to commanding general. The fact that each further step also involves lateral relationships is not appreciated until much later. It is doubtful whether the complex of relationship represented by an institution is clearly grasped before the end of the high school. The appreciation of a complex of ideas behind the institution is even more difficult. It can be only partial at the end of school days.

It is advisable to state again what was said at the beginning of this discussion—the work of test construction was directed chiefly toward those matters which the social sciences had in common,

THE PROBABLE LEARNING PROCESS IN SOCIAL SCIENCE

those to which each contributed more or less directly. The technical knowledge and skill which each of them develop separately have been for the most part disregarded. This arbitrary procedure was justified by the fact that those separate contributions have been most fully emphasized in the vast literature on this problem. The value of those separate contributions is not affected by this emphasis upon the common elements. The skills, the acquaintance with potential sources of information, the technical knowledge and the great reserve of information and ideas which each has to offer are probably only to be acquired in the process of dealing with the materials of each separate field.

One of the conclusions most forcibly gained from this study is that the study of the problem of learning in this field is only just begun. Despite all the effort which has been devoted to the task not only by this Investigation but by the many who have attacked the problem before, we are scarcely over the threshold of the difficulty. Enough has been learned to indicate that the solution of the problem, or, indeed, any real advance toward the solution, will require the best efforts of the social scientists themselves as well as of the scholars in educational measurement and applied psychology. Unless this united attack is made, little real advance will occur, for in the very nature of things the expert in educational psychology can not also advance to expert knowledge in the social sciences; and without that he cannot hope to advance beyond the elementary stages of the problem. The sympathetic and intelligent cooperation of the social scientist will be essential to advance the problem beyond that stage.

CHAPTER II

PLAN AND METHOD OF TEST CONSTRUCTION

The classroom teacher knows that there is a wide difference between what may be offered to the pupil and what that pupil receives or takes away. Parents who concern themselves with the instruction of their children also realize the truth of this statement. Those accustomed to dealing only with adults often forget this fact. An investigation concerned with an educational program for the elementary and secondary school must take it seriously into account.

But the task of discovering how the pupil builds up his knowledge of society is by no means an easy one. The only source of that information is the pupil himself. His testimony alone is conclusive. To obtain his testimony in an efficient way that will yield data useful to those engaged in the construction of curricula and in school instruction presents difficulties. Some form of test through which the pupil will reveal his hazy approach must be employed. But what form of test?

When the Committee which planned this Investigation began its work, faith in the efficacy of the new-type technique for educational measurement was at a very high point. This instrument, which had been developed by psychologists, had achieved widespread attention in the measurement of general intelligence and had been extended with almost equal success to educational measurement. The work of Rice, Ayres, Thorndike, and their followers in the measurement of elementary school subjects had become almost universally accepted.¹ Important progress was being made in the measurement of more advanced subjects of high school and even college grades. The Modern Foreign Language Study, which was just completing its work, had been peculiarly fortunate in its

¹It is impossible here to accord due credit to the many scholars who, as professors of psychology or of educational research, have made significant contributions in the measurement of instruction. The list of those devoting study to specific school subjects includes such workers as Buckingham, Courtis, Dearborn, Gates, C. T. and W. S. Gray, Haggerty, Horn, Judd, Trabue and Van Wagenen as well as many others.

PLAN AND METHOD OF TEST CONSTRUCTION

experience with this new instrument. With the assistance of Henmon, Wood, and Buswell, tests had been constructed to measure progress in the knowledge of vocabulary, grammar, and reading. Specialists in modern foreign language teaching had been skeptical, if not actually hostile, to the use of the device, but those most intimately associated with the investigation were enthusiastic in their praise of the results obtained through the use of the tests.

In view of the great advance that had been made in educational measurement and the high endorsement given to that work by the subject matter specialists in modern foreign languages, the committee planning the Investigation of the Social Studies was led to hope that this instrument might also prove very useful in their work. As yet, few of the subject matter specialists had taken any great part in the construction of such tests. The chief burden for the development of the tests still rested upon the psychologists. The latter had been required to do nearly all of the work on tests designed to measure educational advance at the secondary and even college levels. In exceptional cases, such as that of Henmon, who was himself a highly accomplished scholar in modern foreign languages, and of Wood, who studied law in preparation for constructing the tests for the Columbia Law School, the tests represented a combination of knowledge of subject matter as well as of psychological technique. In general, however, the work of test-construction was still largely in the hands of psychologists whose knowledge of the special subject matter fields was limited. Whether the higher degree of objectivity resulting from a lack of special knowledge would be sufficient to overcome the deficiencies in that knowledge remained still to be determined. Though there was some skepticism among them, nevertheless the planning committee felt justified in providing for a somewhat extensive testing program.

It was the hope of this committee that its testing program might yield instruments by which the relative effectiveness of various curricula and various teaching procedures in gaining desired results might be determined. The great activity in curriculum construction since 1916 had led to the development of many new or different arrangements of courses, subjects, and even materials in the

TESTS AND MEASUREMENTS

field of social studies.¹ While the greatest amount of experimentation had been centered on the three grades of the new administrative unit, the junior high school, there was enough of such experimentation in the elementary, senior high, and junior college grades to afford profitable comparative material. Should the results of the test program reveal decisive advantages in any particular program, or method of procedure, or in both, such results would afford a sounder basis for any recommendations the Commission might make.

¹A COMPOSITE OF SOME SOCIAL STUDIES OFFERED IN

GRADES				
4th	5th	6th	7th	8th
Local Hist. and W. Hist. Stories	Eur. Background and U. S. thru Rev'n.	U. S. Hist., Revolution to present	Eur. Background of American	U. S. Hist.
Local and State History	U. S. Hist. Stories and State History	Eur. Background and State History	U. S. History	Early European
Lessons in Civic Habits	U. S. Hero Stories and Eur. Background	Eur. Background	U. S. History to Civil War	U. S., Civil War to Present
Local Geog. and World Geography	N. Am. Geography and U. S. Geography	West Hem. Geog. and Eur. Geography	U. S. Revolution and Econ. Hist. of U. S.	Civ. Gov. - Local and Special Phases of U. S. History
Local Geography and State & N. Am. Geog.	Civic Virtues	State Geography and Voca'l Opportunities	Eur. Background and Local Civics	U. S., Rev. to Pres and Civ. Government
World Geography and N. Amer. Geog.	West. Hemis. and N. Amer. Geog. Eur. Geography and Asia and Africa Local Civics	Asia, Africa, Australia, and Europe South America and United States Community Civics and Civics Local and State Government	World History Econ. History World Community and Voca'l Civics National Gov't and State Gov't U. S. Geog. and Dependencies and Asia, Africa, Australia State Geography	Community Civics State and National Gov't Community Civics and State Government Math. and Phys. Geog. and Polit. and Comm'l Geog. Tr. Routes, U. S. Geography and Polit. Geography

DETERMINATION OF OBJECTIVES FOR TEST PURPOSES

DETERMINATION OF OBJECTIVES FOR TEST PURPOSES

The Committee realized, however, that hope and the achievement of the task were not identical. It recognized some of the difficulties in the way quite clearly before the work was begun. The Investigation involved a number of subjects which were commonly organized in separate academic departments. A relationship between these several fields of specialization was recognized, but the scholars in all of these fields had been so intensively occupied with their own special fields that few of them had had time to consider

ELEMENTARY AND SECONDARY SCHOOLS THROUGHOUT THE COUNTRY

GRADES			
9th	10th	11th	12th
Anc. and Mod. Hist. and Modern History	Ancient History	Modern History	Problems of American Democracy
Modern Europe	U. S. History	U. S. History	American History and Civics
Ancient History	Med. and Mod. Hist.	U. S. History and Civics	American History and Probs. of Am. Democracy
Early Eur. and Vocational	Later Europe	Hist. S. Am. and Far East	General Social Science
Inter. Rel. U. S.	Modern World History	Greek & Rom. Hist. and Social Problems	English History and Pacific Rim History
Early Eur.	Modern Eur.	Economics and Sociology	Economics and Sociology
World History	World History, to Fr. Rev.	World History, Fr. Rev. to Present	American History
Community Civics and Economic Civics	U. S. History and Hist. of Providence and R. I.		Economic History U. S.
Community Civics	Civ. Gov. and Commun. Civ. & Local Hist.		
American Problems			
Civ. Gov. and Indus. Geography			

TESTS AND MEASUREMENTS

COURSES OFFERED IN THE HIGH SCHOOLS IN 1924

Ancient History
Ancient and Medieval History
Medieval History
Medieval and Modern History
Modern History
World History
English History
American History
American History and Civics
Civics
Community Civics
Vocational Civics
Economic Civics
Problems of Democracy
Social Problems
Sociology
Economics

SUBJECTS OFFERED IN THE SECONDARY SCHOOLS AS LISTED IN
THE *Biennial Survey of Education* FOR THE YEAR 1928

World History
Ancient History
Medieval and Modern History
American History
English History
Civil Government, United States
Community Government
Sociology
Economics
Problems of Democracy
Industrial History
Geography
Current History
Latin American History

DETERMINATION OF OBJECTIVES FOR TEST PURPOSES

Local History
Mythology
Banking
Business Management
Foreign Relations

the exact nature of that relationship. For the sake of economy, moreover, it was desirable to narrow the work of testing to those aspects of the social sciences which offered the maximum amount of comparison. An effort was therefore made to state objectives or aims which all members of the Committee could accept as desirable outcomes of instruction in the social sciences. Such a statement might serve as a guide to the work of test construction.

It proved to be no easy matter to arrive at such a statement of objectives or aims acceptable to all of the Committee. It was difficult to state these aims in terms of specific subject matter because such a statement might limit comparison too narrowly to particular subjects or to schools which offered the same subjects in the same grades. While valuable, this procedure would multiply the number of tests beyond any possibility of completion. Furthermore, the ordinary procedure in schools involved frequent tests of the learning of separate subject matter, whereas the Commission, because of its wide membership, was peculiarly well equipped to consider the aims and values which the social sciences had in common. Accordingly, it seemed best to emphasize, for the guidance of test construction, those objectives, aims, and values which the social sciences had in common. This involved a statement in general terms and one which was intended to have validity in all of the social science fields.

After extended thought and discussion, the Committee was able to agree tentatively upon such a list of objectives for guidance in test construction. This list was included in its working drawings presented in 1928.

1. Understanding of important institutions by means of which society functions. Principles and ideals are included.

This is the most generally accepted function of the social studies, and the most precise. Understanding and information are not synonymous; and tests

Intellectual Aspects	Outlook: approach to novel situations	Key of symbols to be entered in each cell of this table to indicate type of test considered of service for the field in question.										
Orientation:	Factual memory and organized knowledge	In time: one dimensional	(Do not exceed 10 measures)									
			Various historical periods Threads through each from preceding to subsequent	Discipline								
		Scale: Little and inaccurate information — much and accurate (Do not exceed 10 measures)	Geographic Historic interrelationships									
In space: two dimensional		Various geographical divisions	Skill in use of sources									
		Geological, ethnological and economic bonds	Scale: Little and inaccurate — much and accurate (Group in less than 10 measures)									
Q = measure resulting from a questionnaire P = measure resulting from a preference test; items on linear scale M = measure resulting from a multiple choice test, items in categories or measure resulting from a matching test; items in categories F = measure resulting from a free association test G = measure resulting from a map or geography test Ra = measure resulting from ratings of acquaintances O = measure resulting from orientation test. Letters T for time, S for space, C for category following designate dimension involved Re = measure resulting from a reading ability test												

CHART DRAWN FOR HISTORY AND SOCIAL STUDIES INVESTIGATION NOVEMBER, 1929—TRUMAN L. KELLEY

Each caption is to be subdivided into the various disciplines of mankind. For the purpose of the social studies the following divisions should be of service. For convenience in treating this comprehensive field certain rubrics and sub-classes are here indicated. Throughout sharp categorizations are undesirable.

<i>Political</i>	{ Pol for politics and government War for military: land, sea, air Bus for business, manufacturing, and commerce Exp for exploration and discovery Inv for invention Med for medicine Sci for science	<i>Social and Educational</i>	{ Art for painting and sculpture Lit for literature and poetry Mus for music Rel for religion Spo for sport Tea for teaching and philosophy
<i>Economic</i>			

Certain of these are of great importance under certain captions and of negligible importance under others. Scarcely any apply to the "habit" and "attitude" aspects, only a few to "intellectual outlook" and all quite generally apply to "orientation."

	Home: Scale: disruptive— cohesive uncongenial— congenial unfavorable— favorable to personal development	School: Scale: ditto	Immediate Neighbor- hood: Scale: ditto	Extended Contacts: Scale: ditto	Vocation: Scale: ditto	Organized Society: Scale: ditto	Abstract Ideals: re- lationship to God and the universe Scale: ditto	Self-place- ment in Social Order: Scale: ditto
Habits (action habits)	Gross nervous excitation (Do not exceed 10 tests or measures)							
Attitudes	(Do not exceed 10 measures)							

TESTS AND MEASUREMENTS

principles considered important by three members of the Commission, E. B. Greene, Henry Johnson, and the Director. A test or sub-test could conceivably be constructed for each cell in the Table. (+ = test.)

	PURPOSES				INTERRELATIONSHIPS—t5					
	t1—Continuity	t2—Historical sense	t3—Sense of evidence	t4—Interpretation of present	Social studies seldom found in secondary curriculum	Physical and biological science	Economics	Political science civics, and law	Sociology	Geography
PERIODS										
Ancient Orient Greece Rome										
Medieval										
Modern Europe America Latin America										
Contemporary Events Problems of Democracy Contemporary Civilization										

t1—Continuity: It is here intended to measure the pupil's time sense, his place sense, the two conjointly, and also his sense of continuity of related events which, in addition to sense of continuity in time and place, calls for an apprecia-

THE SELECTION OF ITEMS

tion of the continuity of cultures, disciplines, of institutions and of thought, as, for example, is involved in the understanding of the Roman Catholic church or the continuum representing the development of painting.

t2—*Historical Sense*: It is here intended to measure the thoroughness with which the individual transposes his mental outlook to the time, group, or person historically described; his ability to see a century through the eyes of that period; to grasp a motive through the outlook of the one whose act is involved; and to judge a character with such understanding as the character itself, as well as the observer, may experience.

t3—*Sense of Evidence*: It is here intended to measure the critical skill with which the pupil appraises a bit of evidence coming to him under various conditions and with various defects; to measure the soundness with which he draws deductions and makes generalizations; and to measure the mental standards which he maintains in judging of evidence.

t4—*Interpretation of Present*: This in one sense includes all the other values of history. It is, however, intended to measure here such aspects as are not covered by t1, t2, and t3. Here are to be investigated specifically acts of interpretation of happenings in the lives of the pupils and specific personal modifications in conduct or intellectual outlook.

t5—*Interrelationships*: Whereas t1, t2, t3, and t4 have treated of history as an individual discipline, we here wish to examine such relationships as have been established in the pupil's mind between history and other social studies and activities and other sciences. Personal values and independent points of view involving history and several other phases of life at the same time may be here investigated.

The Commission approved the principles and program presented by Dr. Kelley and appointed an Advisory Committee on Tests to aid in the execution of this work. The Committee consisted of Messrs. F. W. Ballou, Chairman, I. Bowman, H. C. Hill, Ernest Horn, Henry Johnson, A. C. Krey, and Ben D. Wood. In consultation with the Committee projects were assigned to individuals and active work was begun.

THE SELECTION OF ITEMS COMPOSING THE TESTS IN THE SOCIAL STUDIES

Immediately preceding,¹ a general outline is given of the field covered by the social studies in the schools. This is the region wherein it was desired to construct tests. The essential service which it was hoped to render by the measuring devices developed was to enable investigators to ascertain whether the values sought and measured by the tests employed were being attained by a given regimen of instruction. For example, if it be granted that the purpose of instruction in the social studies is to develop good citizens,

¹See p. 122 ff.

TESTS AND MEASUREMENTS

nevertheless this idea is serviceable in connection with school problems only to the degree that it becomes so delimited that some means (judgments of teachers or associates, or tests) gives information as to whether, and to what extent, pupils give evidence of "good citizenship." In so far as good citizenship is not quantitatively or qualitatively measurable by tests or the personal evaluation of associates, it is worse than no concept at all, as it distracts the attention from concern with attainable and cognizable outcomes. The moment it is possible reliably to appraise pupils upon this trait, it becomes possible to determine the efficacy of instruction as stimulating the trait; it becomes possible to choose between several types of instruction, or types of curricula, upon the basis of their excellence as developers of the trait.

The technique for determining the efficacy of a procedure is well established. It is that of the controlled experiment wherein the outcomes in terms of the improvement of pupils of two groups, as similar as possible in every respect except the one being investigated, is determined. This technique will not here be elaborated, but it should be borne in mind that it alone provides the setting wherein objective measurements of pupils can establish their serviceability. The basic purpose of the test construction work of the Test Committee has been to make possible the determination of the rate of growth of a class under a certain tutelage and the comparison of this rate with that under an alternative environment. Unfortunately the time required to construct tests, and other limitations of the work of this Committee, have not permitted thorough-going investigations of alternative types of instruction in the social studies by means of carefully controlled experiments employing the tests developed by the Committee and other tests. These things remain to be done in the future, but it is believed that a number of measuring instruments serviceable for this purpose are here made available.

The general field of the social studies as indicated in the preceding sections must be broken up into subsidiary fields as there indicated, and these into still more subsidiary, before a point is reached wherein specific instruction can function and wherein specific

THE SELECTION OF ITEMS

measurement can be brought to bear. This process of analysis and subdivision is distinctly the province of competent social scientists, exercising judgments based upon the sum total of their experiences.

The analyses made by such authorities differ, and the issue that confronts the maker of a social studies curriculum, or a social studies test, is how to proceed in the face of these differences. The simplest way out would be to disregard the authorities and build upon tradition. The accompanying table may help to make clear the nature of the issue.

SPONSORSHIP			
		P.T. = Pioneer Thinkers	T.A. = Traditional Authority
Way Utilized		The problem of first importance here is how to select these authorities. Having them, the problem of how to get and utilize their judgments becomes very important.	A sampling of present practices in regard to social studies curricula, and of the present needs of a citizen to apprehend language, written or spoken, etc., defines this traditional authority. What constitutes a fair sampling is the point of importance here.
Purpose and Content	Nuclear	P.T.N.=Pioneer Thinkers Nuclear outcome. A small nuclear consensus.	T.A.N. = Traditional Authority Nuclear outcome. A large nucleus represented by current practice.
Outcomes of Each Approach	Divergent	P.T.D.=Pioneer Thinkers Divergent outcomes. Wide individual differences as to purpose and content.	T.A.D. = Traditional Authority Divergent outcomes Occasional variations, each the outgrowth of some fraction of tradition.

It is conceivable that the objectives discovered from either the P.T. or the T.A. approach might almost completely overlap each other, but hardly conceivable that the emphasis would be similar. Pioneer Thinkers tend to concern themselves with those matters in the social scheme which require modification and, further, to emphasize such matters in order to gain the desired attention. It

TESTS AND MEASUREMENTS

is to be expected that any selection of objectives based upon pioneer thinkers alone would result in incomplete and distorted views and values, just as it is to be expected that any determination limited by traditional authority would fail adequately to represent society in its growing aspects.

We can readily find fault with each of the four sorts (P.T.N., P.T.D., T.A.N., T.A.D.) of resulting material if used alone for the determination of curricula and of measuring instruments. In brief:

Traditional Authority Nuclear Outcome: The nuclear material of tradition is static and insufficient in a changing social order.

Traditional Authority Divergent Outcomes: The divergent purposes and practices bequeathed by the past may be so numerous as to require selection before they can serve in an actual measurement or teaching situation. In this case the problem of selection calls for "authorities" and the situation becomes indistinguishable from one admittedly based upon the convictions of experts.

Pioneer Thinkers Nuclear Outcome: The nucleus found as a consensus of "best minds" is automatically made smaller and smaller in amount as the number of authorities is increased. It thus becomes most inadequate when the richness represented by the authorities is most complete. Even with a dozen or more authorities, the substance of a consensus may become almost trivial in comparison with the richness of thought of any one of the separate authorities. This consensus may be considered satisfactory as far as it goes, but should not circumscribe curricula or measurement efforts.

Pioneer Thinkers Divergent Outcomes: The divergent views of pioneer thinkers automatically become greater and greater in number as the number of authorities is increased. This approach leads to an unwieldy mass of material. Within this material, however, may be expected to lie the paths of real progress.

THE SELECTION OF ITEMS

If we now endeavor to allocate Test Committee projects to some one of the four sponsoring fields mentioned, our result would be as follows:

Traditional Authority Nuclear Outcome: Measurement of understanding of social science terms projects by Edgar B. Wesley, Mary Kelty and Nelle E. Moore, Mary Gold, and Luella Cole Pressey.¹ Also measurement of informational aspects of history as initiated, but not completed, due to Committee limitations, by David P. Harry, Olive Bucks, and Janet Cameron.

Traditional Authority Divergent Outcome: Nothing.

Pioneer Thinkers Nuclear Outcome: Measurement of social attitudes toward specific features of society as carried out by L. L. Thurstone. The measurement of character traits as undertaken by Truman L. Kelley, M. R. Trabue, and A. M. Jordan.² (Perhaps a minor portion of the above should be classified under P.T.D.) Measurement of fundamental concepts to be derived from a study of geography (part of test material devised by R. D. Calkins and Edith Parker).³ Measurement of certain general outcomes of instruction in history (part of test material devised by Marion Clark).³

Pioneer Thinkers Divergent Outcome: Measurement of geography concepts (part of test material devised by Calkins and Parker). Measurement of general outcomes of instruction in history (part of test material devised by Marion Clark).

It will be obvious that such tests as have been devised are entirely inadequate to represent the divergent views of careful students of the social studies (P.T.D.), wherein one may expect to find, along with much abortive material, the real suggestions for social progress. Of the tests developed by the Committee the one which is most inadequate as a measure of tradition is the Calkins-Parker Geography Test, or at least the major portion of it, while it is at the same time the most promising one developed as a measure of what may prove to be an important and real line of development in future work in geography.

¹See Chap. III.

²See Chap. V.

³See Chap. IV.

TESTS AND MEASUREMENTS

A quite radically different technique is appropriate in the construction of tests which are the outgrowth of sampling traditional practices (T.A.) from that appropriate to tests sampling the ideas of pioneer thinkers (P.T.). When tradition or current practice constitutes the warrant for the inclusion of material in the curriculum, it is sufficient to ascertain what current usage is, and construct test items in harmony therewith. The resulting test scores measure the extent to which a child is aware of the prosaic contemporary social life about him. If a word such as "financier" is in common usage in the social studies and in newspapers and magazines of wide circulation, it calls for the judgment of no social scientist of renown to establish that knowledge of the word is to a degree (of course slight, as but one word is in question) indicative of a higher level of understanding than lack of knowledge of it, and that such knowledge is in fact "good." In the dim past some pioneer thinker introduced the term, but that is long forgotten and today it is merely a matter of noting that the word serves in understanding life about us to conclude that development of knowledge of it is a desirable feature of a curriculum, and that measurement of understanding of the word is an informative service that can be rendered by a test. Clearly all common concepts are of this nature and this is irrespective of any emotional tone accompanying them. The arch-conservative could hardly deny that to know the meaning of the now common word "communist" indicates a higher level of social understanding than not to know it, and thus "communist" becomes a fit word for inclusion in a social concepts test irrespective of sponsorship by eminent social scientists.

With these introductory remarks we may now proceed to a statement of the specific steps to be followed in constructing a test of knowledge of terms used in the social studies. The selection of terms to be incorporated into the test comes first. These terms are to be drawn from a very great number—if possible, from a number so great as to represent practically the entire field. The

THE SELECTION OF ITEMS

steps actually followed in obtaining such an initial list have been described in connection with the various tests developed to measure knowledge of social terms. The list of social terms is so long that it obviously is not feasible to incorporate all of them into the tests devised. A sampling of these terms is resorted to, and is dependent upon the judgment of the person actually constructing test exercises. Because this person must exercise judgment in the selection from the larger list of the words from which he will construct test exercises, and because his judgment is involved in the actual construction of each test item, his work, but not that of those making the initial word counts, is to be subjected to the scrutiny of other students of the social sciences.

The form of test item that he employs may be investigated experimentally by trying a number of forms and choosing the one or two yielding the most reliable results (as was done in the case of the Gold tests), or it may be decided upon from general considerations built up from a knowledge of current tests in related fields (as in the main was the case with the other tests developed of social terms). The result of both of these approaches was an adoption of the multiple-choice test with four or five options.

The initial efforts at item construction were scrutinized by the psychologist of the Commission for errors in form. Things that were especially noted were:

(a) Accuracy as determined by reference to a standard dictionary, not as judged by a teacher of the social sciences.

(b) Fitness of balance between the word difficulty of the essential word being tested for and the words ancillary to the process of testing. For example, to employ the word "indescribable" (having a frequency value of 5, i. e., in the eighth thousand in the Thorndike *Word Book*) when testing for knowledge of "church" (frequency value of 81, i. e., in the fifth hundred in the Thorndike *Word Book*) would pervert the test, it becoming a test of knowledge of "indescribable" and not of "church."

TESTS AND MEASUREMENTS

(c) Lack of secondary response cues. For example, if the phrase representing the correct response option is generally longer than those representing the alternative options, then the length of option becomes a cue to response. There are many secondary cues: an order of response option other than chance may be one; special type of qualification of option, as, for example, those employing the words "never," "always," "greatest," "least," and ultimates of any sort, may commonly characterize incorrect responses. Grammatical inconsistency may be such a cue, for ordinarily the correct option is euphonically and grammatically in agreement with the initial portion of the exercise, but "fillers" merely stuck into to provide options are less certain to be.

(d) The order of difficulty of test items in preliminary forms of the test. It is a dogma of good test practice, easily defended, that the items of a test should be such, and so arranged, that the test starts with the easier items and continues by as regular steps as possible to the harder items.

(e) Fairness with which the field was sampled. Though this point was at times considered by the psychologist, it was mainly left to the test author and the social science critics.

The things mentioned all occur early in the process of item scrutiny. They were followed by a review of items by social science specialists to insure essential accuracy in the technical aspects of the subject.

I would here mention some failures in procedure. This is done at the risk of seeming to be unsympathetic with the trials of the social science critics, but it is in truth done with the view to pointing out the serious difficulties confronting one who attempts to use high grade talent in criticizing elementary school items. Most words possess a hierarchy of meanings, as, for example, the word "railroad." Beginning with the local railroad station and train as evidence of primitive meaning, one may continue to systems, mergers, Interstate Commerce Commission, as elements in a fairly full mean-

THE SELECTION OF ITEMS

ing. For a child to get credit for saying that a railroad consists of tracks with cars upon them, may be considered ludicrous by the university professor of transportation. However, this may represent a degree of knowledge quite accurately characterizing first grade understanding, and not kindergarten understanding. In other words, the item may have a legitimate function as a primary test item. The authorities, generally university men, who so kindly cooperated with the Commission in judging items, proved to be very severe critics, tending to eliminate certain very sound items merely because of a difference in the conceptual level in the mind of the deviser of the item from that in the mind of the critic.

Further, an allied outcome of the judgments of items was the general severity of destructive criticism unrelieved by constructive suggestions. Had it been required that each adverse criticism be accompanied by a constructive modification of the item criticized, the work of the devisers of test items would have been greatly simplified. The observation that an item is inadequate, for some very general reason such as that "it misses the point," is not especially helpful. Under the conditions existing, which were that the judges donated their services, to ask for constructive criticisms would have been unreasonable, but it would not be unreasonable if services were adequately paid for. The point of this discussion is in no sense to find fault with the judges of the tests or with the manner in which they worked. It is rather to show the inherent difficulties involved, in order that future endeavors may not be mired or made difficult by the same problems. The psychologist cannot but feel the highest esteem for certain authorities, whose lives are dedicated to college and graduate school instruction, for the consideration they showed when called upon to judge items designed to measure elementary and high school levels of ability. The judgments rendered have resulted in an improvement in the elementary and high school tests, but at the cost of annoyance to the judge and to the deviser of the test items (for all complaints ultimately reach him),

TESTS AND MEASUREMENTS

that it is desirable to avoid. The essential suggestion offered to avoid the difficulty noted is that judges be adequately paid and asked to accompany each adverse criticism with a constructive suggestion. It is not suggested that judgments of experts of professorial grade are not needed and should not be called for in connection with the construction of elementary test items in the field of the social studies.

Having received the criticisms of the judges, the author of the test then revised the items, and the process of criticism was repeated with the idea that finally a test would be constructed that would satisfy both the author and his critics. Generally speaking, this was accomplished, though it should be noted that there was more or less reservation, though not positive disapproval, on the part of the judges as to the value and fitness of a number of the test items, even of some of those in the final forms.

The items were then arranged, as nearly as possible, on the basis of the author's judgment, in order of difficulty and built into preliminary forms for try-out. These were given under what were intended to be standard directions to the appropriate grade groups dispersed as widely throughout the country as was feasible. All papers were sent to the psychologist for scoring and analytical study of the performance of test items (except in the case of Dr. Pressey's test, which was scored under her direction).

Information yielded by the try-out. By behavior of an item is meant the change in the percentages of correct responses to the item from groups of various levels of ability. It is thus necessary to have groups of different levels of ability in the function which is being tested for. If a test of "introspection" is being devised, and a test item thought to measure it (perhaps actually measuring it) is given to various subjects, unless there is an independent means of segregating the subjects on the basis of "introspection," there is no way of usefully describing the behavior of the item. To say that 40 per cent of group A pass, and 60 per cent of group

THE SELECTION OF ITEMS

B do so, means little if it is unknown whether or not group B is superior in "introspection" to group A.

Three ways of determining groups differing in ability with respect to the function being measured have been followed and are mentioned in order of decreasing merit.

(a) Judgments of teachers, or of other associates, have been utilized to select subjects (pupils) having different abilities.

(b) Age or school grade location has been taken as a basis for selecting groups, on the assumption that the higher the grade, the higher the ability. This obviously will work well for traditional school values and purposes, for here grade location has been determined by judgments of teachers wholly or in part upon the ability of pupils in the mental function in question. This method has been largely followed in connection with the study of knowledge of social science terms where increase in age or school grade may be assumed at least to go with increased knowledge. This method was used, though not with complete satisfaction, undoubtedly because of the nature of the functions being tested, in studying the Marion Clark historical ability items, and the Calkins-Parker Geography Test items. The method was not even investigated in connection with the Kelley-Trabue character measures or the Thurstone social attitudes scales, for in these cases the assumption that pupils in higher grades stand higher in the functions involved is clearly not established.

(c) The performance of pupils upon the entire preliminary test has been taken as giving an order of ability in the function which could be utilized in studying the performance of the separate items successively.

In a very important way this last procedure begs the question. Strictly speaking, all it can do is to reveal whether the single item measures the same thing as the average of all the preliminary test items. If the large number of items composing the preliminary test do, in fact, measure the function they are claimed to measure, ex-

TESTS AND MEASUREMENTS

cept for chance errors, then the average of them all (characterized, as an average is, by greater reliability than the separate measures entering into it) does constitute a sound criterion whereby to judge the separate items. The method therefore does have the merit of lessening the role of the chance factors entering into the instrument of measurement. It has no merit whatever in eliminating systematic errors represented by the difference between the average function measured by the items of the preliminary form and the concept as warranted by the title of the test. When this third method of studying the performance of items is resorted to, the full weight of responsibility for the validity of the test rests upon the authors, for this validity is in no wise affected by the results of the try-out. In the Calkins-Parker Geography Test, for instance, the try-out quite generally failed to reveal important grade differences. Accordingly, neglecting grade grouping (supposed at the beginning of the investigation to indicate differences in levels of geography ability) we have left evidence of validity, which goes back to the authors of the test items, but not back to differences of performance of pupils known to stand high and others known to stand low in the ability in question.

Though study of the performance of an item in groups scoring successively higher upon the total preliminary test provides no basis for keeping or rejecting the item, because it does not measure the function called for in the label of the test, the performance of an item does provide a very good basis for selection for the excellence with which an item measures whatever it is that the average of the remaining items in the try-out form measures. Items selected in this manner can be built into a final test which does consistently measure something, but whether it is what the test label calls for remains a matter of judgment.

In the case of material for which it was assumed there would be progress from school grade to grade, item profiles were constructed, as, for example, the following taken from the Kelty-Moore Test of Concepts in the Social Studies:

THE SELECTION OF ITEMS

Item A: The terms of a treaty are
the length of the treaty
the matters agreed upon in the treaty
each half of the treaty
how long it took to make the treaty
the places where the names are signed

Item A. *Profile of a Good Item*

Per cent correct	Grades					
	4	5	6	7	8	9
100.....	98 (9a)
90.....	91
80.....	82	85	84
70.....
60.....	56
50.....
40.....
30.....	30 (4a)	27
20.....	20
10.....	10 (4b)
0.....

TESTS AND MEASUREMENTS

Item B: Staples are

things which always remain steady

places where horses are kept

the principal products used in trade and commerce

trees from which syrup is obtained

posts in the ground

Item B. *Profile of a Poor Item*

Per cent correct	4	5	6	Grades 7	8	9
100.....
90.....
80.....
70.....
60.....
50.....
40.....	38 (9a)
30.....	33
						28 (9b)
20.....	19	
	14 (4a)	16		11		
10.....	9	..	10
	4 (4b)					
0.....

THE SELECTION OF ITEMS

Item C: To subdue means
to stop suddenly
to conquer by force
to pay money you owe
overdue
to give in to a greater power

Item C. *Profile of a Useless Item*

Per cent correct	Grades					
	4	5	6	7	8	9
100.....
90.....
80.....
70.....
60.....
50.....
40.....
30.....	28 (9a)
				25	25	24
20.....	..	19	20 (9b)
			11			
10.....	10 (4a)
	9					
	8 (4b)					
0.....

TESTS AND MEASUREMENTS

An item such as *A* has every indication of being a good measuring instrument. It is good in the sense of reliably measuring growth from grade to grade in whatever it is that is involved. One final check should be made to provide evidence as to whether the exercise continues to be a measuring instrument below and above the mean 4th grade and 9th grade levels. To be serviceable, the item should differentiate for these two terminal classes as well as for the intermediate classes between pupils low and high in the ability being measured. We may therefore divide the 4th grade into two sub-groups, those below the median upon the total preliminary test score and those above it. To be a good item at the 4th grade level the first of these groups should do less well than the second. Similarly for each other grade, but it is hardly necessary to go to the trouble of investigation for the intermediate classes since evidence is available for grades upon either side. Accordingly we will do this for the two extreme classes only, the 4th and the 9th. The records of the lower and upper halves of the 4th and 9th grade pupils are indicated on the profile charts opposite the symbols (4b), (4a), (9b), and (9a). Since (4b) is decidedly below (4a), and (9b) below (9a), and since from the mean 4th grade to the mean 9th grade level inclusive there is definite and substantial progress in the percentage of correct scores, the item meets every statistical requirement of a good item. The study here suggested of the performance of the upper and lower halves of terminal classes is to be recommended, but it was not generally pursued in the actual study of the items in the social studies tests, so that the tests as built up are probably not quite as reliable for terminal classes as for intermediate classes.

Item *A*, having earlier met the requirements of the judges of the items as to substantial accuracy, and being known from word counts to involve a concept commonly entering into the social studies, may be considered entirely satisfactory for measurement purposes.

Items *B* and *C* are not satisfactory from the standpoint of their behavior. The study of the performances of items has the purpose of building up a test from such items as *A*, rather than from such

GRADE RELIABILITY CRITERION

as *B* and *C*. Items of the *A* sort will be referred to as "good" items, of the *B* sort as "poor" items, or as items having little diagnostic power, and items of the *C* sort as "unsatisfactory" items, or items having no diagnostic power, meaning thereby no power to diagnose any ability of pupils as represented by differences in the mean accomplishments of successive school grades.

GRADE RELIABILITY CRITERION

In case there is little or no progress in the function in question from grade to grade (as may be the case in, say, cooperativeness), but real differences between pupils within any grade, the profile of the item would be as shown in *D*, where (5 b), (5a), etc. refer to lower and upper halves of the classes in question on the total score of the preliminary test in question.

Item D. *Profile of a Reliable Item Which is Useless in Diagnosing Grade Differences.* (Hypothetical)

Per cent correct	Grades					
	4	5	6	7	8	9
100.....
90.....
80.....
70.....	70 (4a)	..	69 (6a)	71 (7a)	70 (8a)	70 (9a)
60.....	..	68 (5a)
50.....	50	49	51	50	49	51
40.....	33 (6b)
30.....	30 (4b)	30 (5b)	..	29 (7b)	28 (8b)	32 (9b)
20.....
10.....
0.....

TESTS AND MEASUREMENTS

Here is an item which definitely distinguishes between the high and the low in the function in question. It shows no systematic grade differences, presumably because the function in question is not growing from grade to grade. The statistical technique for studying the items entering into such a function is thus, of course, quite different from that for studying the sort previously mentioned. For studying this function we need to know the reliability of the item, or of a sampling of similar items, for each grade. In the case of the Calkins-Parker Geography Test, the first endeavor was to secure evidence as to good items by studying their grade profiles. These very generally showed little progress from grade to grade, so it was thought that the items and the function involved might be of the *D* sort. Accordingly the reliability of the items was investigated by correlating the score on a certain number in one category of the test with an equal number of the same category. However, reliability appreciably above zero was not found for many of the geography items which had been most carefully devised to measure certain important geographical abilities. With reference to these we are thus forced to conclude that the items are not measuring anything as yet existing in definite and different amounts in the school children of the classes studied.

A number of these items, however, did show reliability with groups of college freshmen, so that evidence is at hand that they are measuring something, provided that something exists in the pupils examined. We may think of these items as probably being ready to measure junior and senior high school pupils in a certain sort of geographical ability as soon as such pupils have been sufficiently exposed to the ideas inherent in this ability to show and develop individual differences. The test has thus not established itself as a serviceable measuring instrument, but for those geographers who trust the ability of the sponsors of this test to define important geographical concepts it is a challenge so to instruct pupils that the ability in question does come into existence. This is a very important challenge and it has been brought to light by the attempts to measure an important function. This fact should be noted by such critics of test work as maintain that no basis for

VALIDITY CRITERION

“progress” is to be found in the measurement movement. On the contrary, the movement teems with many suggestions to one interested in measuring intellect in whatever manner it may be conceived or may show itself, rather than in measuring some predefined and circumscribed ability.

VALIDITY CRITERION

We may consider increase in percentage of correct responses from lower to higher school grades as evidence of some ability which does progress from grade to grade, but it is not evidence as to what this thing is, whether, for example, it is spelling ability, weight, computation ability, or something else. Having an item yielding the profile *A*, we are not sure whether it is a spelling item, a measurement of weight, or what not. What it is has been assured to us only by the judgments of the author of this test and his advisers. Now it sometimes happens that such judgments may be very inaccurate. For example, an item intended to measure honesty might in fact measure dishonesty; that is, a high score anticipated to be indicative of honesty might be indicative of the opposite, just because the deviser of the item was not sufficiently expert to appraise correctly all the psychological elements determining the ways and the reasons of a pupil answering the item. In this case it is quite unsound to construct test items on the basis of judgments. (An example rapidly becoming classic is with reference to introversion-extroversion where substantially as many different sorts of functions have been claimed to be introversion-extroversion as the number of tests of it built up by different authors.) Here the appropriate procedure is for the judges to select subjects (pupils) standing low, and others high, in the trait, but not to select or construct test items, and then experimentally discover test items which segregate the subjects in the same manner as they have been segregated by the judges. Such items, no matter what their content, or how peculiar they may “look” to the critic, are actually measuring the function in question. This is the method that has been followed in the case of the Kelley-Trabue character measures.

TESTS AND MEASUREMENTS

We may illustrate by one of the eight traits which it was attempted to measure. It is conceivable that exercises, perhaps somewhat like the sample below, might be built up to measure cooperativeness of pupils.

Item E, for a fourth grade group:

Four children were building with lumber and blocks in the yard. Each made a remark. The first said,

"Let me have my share of the blocks and I will build in the corner."

The second said,

"Let us all build together and make a house."

The third said,

"I want to build a great big tower."

The fourth said,

"When you get through building I will be a hurricane and knock down what you have built."

Which of the four remarks is most like a remark that you yourself might have said? Put a check mark in front of it.

Our adult judgment might assert that to check the second is evidence of cooperativeness. Perhaps this is correct, but the construction of appropriate exercises is extremely difficult, and the items are likely to reveal the intent of the test and thus vitiate it. There are so many pitfalls that it is hazardous to trust one's judgment in this matter. On the other hand, if a school teacher knows the four children in question, he can with considerable certainty rank them for their general spirit of cooperativeness. Judgment is here upon fairly trustworthy ground and represents a sort of average opinion based upon various contacts with the pupils. Judgment of an item, however, is very likely to be a judgment in vacuo, the item not having come within the direct experience of the person judging. This, for example, would be the case if a teacher had never observed the conduct of enough children who had said, "I want to build a big tower," to know whether it was a cooperative or a non-cooperative attitude that was being manifested.

In fields not characterized by a substantial and recorded traditional content the *a priori* construction of test items offers such difficulties and uncertainties that it is doubtful whether it should be attempted; or, if attempted, it should be non-determinative of final test construction outcomes unless verified by an independent

VALIDITY CRITERION

study of validity. It was not attempted in the process of securing measures somewhat similar to cooperativeness, such as courtesy, fair play, honesty, but a procedure considered to be much sounder was followed. Teachers were asked to rank pupils for courtesy. Then a free association test (wherein a child is called upon to respond with the first word coming to mind after seeing and hearing a stimulus word) was given. The responses were then classified according to the judged traits of the pupils making the responses. If the stimulus word is "school" and a number respond with "mates," and if the average courtesy ranking of such is distinguishably (*i. e.*, sufficiently so that chance is a probable explanation) above the average of pupils in general, there is indication that to respond with the word "mates" is indicative of courtesy. The weaknesses of this method of determining what are courteous responses are the size of the chance factors involved, the laboriousness of determining the significance of each of the very large number of responses made, and, finally, the laboriousness and cost of scoring the responses of a newly tested child.

This is not the place to discuss these difficulties in detail or to suggest how each might be met. It here suffices to point out that a very different criterion is being employed, namely, judgments of competent acquaintances of individuals, not judgments of test items. With this procedure there can be no question (granting computational accuracy) that the trait being measured by the responses as scored is the trait indicated by the test label—courtesy. Certainly courtesy means that trait in individuals that is so characterized by competent acquaintances, and nothing more and nothing less. Accordingly in the measurement of character traits by means of the free association test here undertaken, there was no call for, nor dependence upon, judgments of social scientists operating upon test items, but there is utter dependence upon the appraisal of traits of individuals made by their acquaintances, *i. e.*, teachers and fellow pupils.

In summary: Tests of the social concepts ~~sort~~ are considered valid because (1) current usage provided the terms, (2) social scientists certified to the substantial accuracy of the distinctions,

TESTS AND MEASUREMENTS

or meanings, found in each test item, and (3) average progress from grade to grade was substantial.

Tests of historical (Marion Clark), and geographical (R. D. Calkins and Edith Parker) abilities, and of attitudes (L. L. Thurstone), are considered valid because (1) of the competence of the social scientists defining the functions tested, and (2) of trust in the ability of test authors to devise exercises measuring these functions.

Tests of character traits (T. L. Kelley and M. R. Trabue)¹ are considered valid because (1) those possessing a trait are, except for chance errors of judgment, those thought by acquaintances to possess it, and (2) the scoring scheme built up is such that the test reveals, except for chance errors, just the trait differences asserted by the acquaintances.

This covers the types of criteria used in the construction of the various sorts of social studies tests. A more detailed illustration of their use is to be found in connection with the separate tests.

¹See Chap. V.

CHAPTER III

TESTING OF UNDERSTANDING

The list of objectives drawn up by the planning committee described the objective of understanding as follows:

“Understanding of important institutions by means of which society functions. Principles and ideals are included.

This is the most generally accepted function of the social studies, and the most precise. Understanding and information are not synonymous; and tests should seek for *understanding*. To test this objective effectively, it will be necessary to break it up into parts.

Political institutions—local, national, international; principles and ideals.

Economic institutions, local and general; principles and ideals.

Social institutions, domestic, ethical and religious; principles and ideals.

Educational institutions—principles and ideals.

Esthetic institutions—principles and ideals.

Recreational institutions—principles and ideals.

These institutions, principles, and ideals, are to be drawn from the standard text-books and syllabi at research centers.

The material thus collected is to be used as the basis for the construction of tests.

In 1929, when the work on test construction began, it seemed highly desirable to break up this objective as minutely as possible, into single words or at the most very brief phrases. The limitations which this procedure involved are discussed elsewhere.¹ To discover single words which might be used to test understanding rather than memoriter information suggested some distinction. The pioneer work of Van Wagenen, the work of Woody and Steph-

¹See Chapter I., pp. 5-18.

TESTS AND MEASUREMENTS

enson, of Mrs. Pressey and of Harold Rugg's students was studied for possible clues. Finally it was decided to distinguish terms designating direct relationship from those which designated persons, objects, events, and circumstances related. Most proper nouns and words were thus eliminated and attention was concentrated chiefly upon common nouns and terms. A list of such terms was compiled from standard textbooks used in the social studies through the junior college years.¹

The authors of the tests of *understanding* constructed on this basis for the Commission have described their work in the following essays. Wesley and Gold worked at the secondary school level, Wesley continuing the work on the junior college level. Kelty constructed tests for the intermediate grades which could also be used at the junior high school level. These three worked immediately under the direction of the adviser on tests. All are teachers of the subject-matter. As a check upon their work, Mrs. Pressey, a psychologist long interested in this problem as one of educational measurement, constructed her tests independently. The comparison of results was to prove interesting and instructive.²

¹See Appendix I for this list.

²See Chapter I, p. 37 ff.

A STUDY IN THE LEARNING OF THE FUNDAMENTAL SPECIAL VOCABULARY OF HISTORY FROM THE FOURTH THROUGH THE TWELFTH GRADES

LUELLA COLE PRESSEY
Ohio State University

I. PREVIOUS RESEARCH USED AS A BASIS FOR THE PRESENT STUDY

For about ten years the writer has been in charge of remedial work done with college students who were on probation because of academic failures. From the beginning of the work it became obvious that many of these probation students had fallen into scholastic disfavor because they could not read their textbooks intelligently. This problem of reading was especially acute in the social sciences. In analyzing the difficulties involved, the writer inevitably and quickly came upon the paucity of vocabulary displayed by these students as one of the major causes of inadequate comprehension of what was read. It seemed evident then that a step in the remedial program would be the building up of a more adequate vocabulary. However, even a casual glance at textbooks in the social sciences revealed literally hundreds of special words, quite evidently varying greatly in frequency of appearance, importance, and difficulty. The task of teaching the meanings of all such words to inefficient readers of social science texts was obviously both impossible and unnecessary. The writer therefore set about finding out which were the essential technical words with the idea of teaching these first. This research turned out to be a real task, requiring some eight years for even the solution arrived at to date.

In determining which were the really essential words the writer analyzed first twenty-three textbooks in history and civics used in grades from the fourth through college, and listed what words occurred, without noting the frequency of each word. Later she

TESTS AND MEASUREMENTS

analyzed six history texts widely used in the high schools, this time keeping a count of how many times each special word occurred. Another step was to make use of published results of other similar analyses in history and civics.¹ Finally, frequency counts were made of the special words occurring on the front pages of newspapers, in the editorials, and in many articles in magazines dealing with current events.

From a combination of word lists culled from these various sources a master list of 1,444 *different* words was constructed, the frequency of each word appearing after it on the list. The matter of mere frequency of appearance was one criterion used in selecting the essential words.

In addition, sixty-four teachers of secondary school history and five professors of college history rated each of the 1,444 words. They marked each word in one of three ways: "essential," if they thought they could not teach without it; "accessory," if they regarded the word important but not essential; or "unimportant," if they thought they could get along without it. A combination of these ratings served as the second criterion.

The entire list of words was next rated by seven especially trained individuals² who expressed their feeling as to the "sociological value" of each word—that is, its use outside the history classroom. These ratings constituted the third criterion. As a final step, the writer went over the total results from all three types of investigation—frequency, importance, and sociological value—and selected those words that were frequently used, considered important by teachers, and of high value outside the classroom. Any words not meeting all three criteria were eliminated from this list of absolute essentials. The selected list contained 415 words

¹Barr, A. S., and Gifford, C. W., "The Vocabulary of American History," *The Journal of Educational Research*, XX: 103-121, 1929. Also Stephenson, A. W., "The Special Vocabulary of Civics," *The Journal of Educational Research*, XVIII: 297-304, 1928.

²A college professor whose specialty is professionalized subject-matter in history, one college professor of history, an experienced and intelligent social worker, a mature high school principal who had taught history for several years, a public school specialist in educational research, a specialist in test construction, and the writer.

PREVIOUS RESEARCH USED

(of which, however, only 346 were actually tested, the remaining 69 being omitted for one reason or another, used as examples, or appearing in unsatisfactory test items). The list may be found in Appendix A at the end of the paper.

The next step was to construct a test by means of which comprehension of these minimal terms might be measured and specific shortcomings discovered¹; so that remedial work might be done. An objective form was used because of its high reliability¹ and the great convenience in its administration. Special effort was made in the construction of the test questions, to avoid mere definitions and to present each special word in a contextual setting that would not be too unlike the setting in which it might appear in a text. A few examples are presented below:

11. What is sometimes given to people to prove their ownership of the land they settle upon? (a) petition, (b) charter, (c) doctrine, (d) amendment.
53. What happens when the value of money is inflated? (a) it will buy more, (b) it is not possible to invest it, (c) it can be used to pay international debts, (d) it will buy less
68. When is an injunction most often used? (a) during a strike, (b) during a battle, (c) during an expedition, (d) during a voyage
76. When a law is repealed what is always the result? (a) it is enforced more rigidly, (b) it is amended, (c) it is no longer in force, (d) it is a source of dissatisfaction

The work had already been completed up to this point before the grant was made by the American Historical Association for the research reported in the main part of this paper. With the beginning of the research under these auspices² the first step was to

¹Reliability means simply that the scores on one form of a test correlate well with scores on another form; that is, the same pupil tends to make the same score on two or more forms covering the same material. The average coefficient of reliability for single grades containing 160 to 190 pupils of the first 80 items of the test constructed as compared with the second 80 is .89; that between the third 80 items and the last 80 is .90. These coefficients of reliability are so high that there can be no question as to the reliability of the materials. The matter of validity—that is the question as to what the test actually measures—will be returned to later.

²More directly, under the supervision of the Commission on the Investigation of the Social Studies in the Schools.

TESTS AND MEASUREMENTS

have representatives from the various social sciences go over the questions to be used, with the idea of making them as acceptable as possible to teachers in these subjects. The writer's original formulations of 325¹ questions were, therefore, gone over and critized by eight people from the social sciences. The final formulation of each question is more or less a compromise among the various points of view. The writer is quite willing to admit that the wording is not always satisfactory; in fact, three items had to be thrown out of the results because they were too poorly presented even to score. The remaining questions, although not perfect, seemed to be well enough phrased and to have good enough "confusion" answers (with a single, unmistakably right answer) to yield information of interest.

For convenience in handling, the 325 test questions were presented in four four-page folders, with 85 items in the first folder and 80 in each of the other three. These folders will hereafter be referred to as Forms A, B, C, and D. The selection of questions for each form was supposedly entirely random (as was also the order of questions within each form), but as a matter of fact the student who made up the forms did not follow instructions exactly, with the result that an undue proportion of the political terms got into one form, too many of the military into another, and so on. However, the forms are of no separate significance and were used merely for ease in testing.

II. PURPOSE OF THE INVESTIGATION

The chief purpose was to trace the learning of each of these essential words from the low fourth grade through the twelfth. It was also desirable to observe the general trend of the learning of the total vocabulary. When the mastery of each item of special vocabulary had been shown, it was the writer's intention to compare vocabulary achievement at each level in school with analyses of textbooks used, to see what relation there might be between words actually known and words assumed by the texts as known.

¹Because a few questions involved the use of more than one of the special words, there were fewer questions than there were terms.

PURPOSE OF THE INVESTIGATION

A further comparison of words appearing frequently in newspapers and current magazines with mastery at the end of the twelfth grade seemed desirable as a means of estimating how complete the training in school was—provided a pupil finished high school—as a preparation for the reading of current history. There were, then, two major objectives: (1) to determine the progress of mastery through the grades studied, and (2) to compare this mastery with (a) the vocabulary used in the textbooks read at various levels and (b) the vocabulary needed for intelligent reading of newspapers and magazines.

For the purposes just enumerated the materials are subject to certain obvious criticisms which the writer prefers to consider before taking up the results obtained by the testing. In the first place, (1) the results show only recognition of the meaning of the words and not spontaneous definition. However, "merely recognition" is exactly what happens when one reads; perhaps this element in the situation is an advantage rather than a handicap.

Secondly, (2) the test obviously measures comprehension of each term in the single context presented and does not tell the extent to which these same terms (a) would be recognized in equally good but differently worded test items, (b) would be understood in context when an assignment was being prepared, or (c) would be defined correctly in spontaneous writing. Studies of the agreement of duplicate test items show, however, high correspondence between the duplicates; thus it may be supposed that the percentage of children correctly answering these questions will not vary greatly from the percentage that would respond correctly to another test item of similar objective type.

The question as to the extent to which recognition of word meanings in a test corresponds with similar recognition while reading a textbook is obviously harder to investigate. The writer has here only two studies of her own to offer. In one, the relationship between correct scores on 40 vocabulary items correlated with comprehension of passages containing these vocabulary elements with a coefficient of .79. In the other, the correlation between marks in history (depending presumably to a great extent upon reading)

TESTS AND MEASUREMENTS

and scores on a history vocabulary test (for 387 cases) turned out to be .67.

As regards the relationship between spontaneous definition of terms and recognition of these same terms in objective form the writer has one further study. In history classes 387 students were asked to define 25 words selected from the tests. These responses were graded. About a month later these same students were given the objective tests. These responses were also scored. The results were then summarized in terms of the percentage of agreement between written definitions and recognition. It was found that there was perfect agreement in 70 per cent of the instances; that is, 70 per cent of the time those who answered a given test item correctly also defined the words involved correctly, and those who failed an item were unable to give satisfactory definitions for those particular words.¹ On the whole the writer feels that what evidence there is points to the conclusion that recognition of meanings in one setting is more likely than not to be directly associated with recognition in another setting.

(3) Clearly, a test item can test only one meaning of a term. The writer is quite willing to admit that the particular meaning of a test item that is offered is not always the one most frequently used, not always the best one, and sometimes is a colloquial, rather than a rigidly technical, meaning. Some of the meanings are at a very elementary level; others require far more mature understanding. If anyone else had been mainly responsible for the items, the same situation would, in all probability, have appeared. However, it must be remembered that "comprehension" inevitably means "comprehension as called for by this or that particular question." This limitation is not due entirely to the objective form; it enters into any question asked, whether the answers are supplied or are to be written out spontaneously.

(4) As mentioned earlier, a few items proved to be entirely unsatisfactory; that is, they had no clearly correct answer or else had more than one equally good answer. Results on these items have been omitted.

¹This material is to be published later by Dr. Elinor Barnes and the writer.

PURPOSE OF THE INVESTIGATION

In spite of the shortcomings, readily acknowledged, the writer feels that the results presented are significant and of practical value to teachers.

III. METHODS OF PROCEDURE IN THE INVESTIGATION

The data presented below are based on results from 16,682 tests taken by over 11,000 pupils in the fourth, sixth, eighth, tenth, and twelfth grades of thirty-one different places. There was an average of 4,171 cases per form and 834 cases per grade. The number of cases per form and per grade, as well as the geographical distribution of the places cooperating¹ in the project is shown by Table I.

At this point recognition should be made of one technique used, without which the testing of such a large number of cases with the money available would have been impossible. In the first place only 4,000 blanks were printed.² None of these were marked at all (with the exception of a few used to get comparative data as to expense) and so they could be used over and over again. What

¹The writer is indebted to the following persons for their assistance in gathering the data: Dr. Brian E. Tomlinson of Emporia State Teachers College and the superintendents of Americus and Emporia; Dr. H. H. Bixler of Atlanta, Georgia; Dr. Virgil E. Dickson of Berkeley, California; Mr. W. C. Kennard of Harvard University and the school officials of the Cambridge high schools, and of Wenham, Beverly, and South Hamilton, Massachusetts; Principal Elizabeth Fassig of Columbus, Ohio; Dr. Charles E. Greene of Denver, Colorado; Dr. L. V. Cavins of Charles-town, West Virginia; Dr. Arthur Repp of the State Teachers College at Flagstaff, Arizona; Superintendent S. M. Brownell of Grosse Point, Michigan; Mrs. Mary Lakoff, Supervisor of Social Studies in Hamtramck, Michigan; Superintendent I. M. Allen of Highland Park, Michigan; Dr. W. W. Kemmerer of Houston, Texas; Dr. Anna G. Myers of Kansas City, Missouri; Dr. I. N. Madsen of Lewiston State Normal School, Lewiston, Idaho; Dr. Ernest Branson of Long Beach, California; Dr. Dean A. Worcester of the University of Nebraska and Superintendent Lefler of the Lincoln, Nebraska, schools; Dr. W. W. Coxe of the State Department of Education in New York; Superintendent E. E. Jones of Mechanicsville, New York; Dr. W. W. Theisen of Milwaukee, Wisconsin; Dr. J. J. Forester of Montclair, New Jersey; Superintendent C. G. Hetherington of Penn Yan, New York; Dr. Philip A. Boyer of Philadelphia, Pennsylvania; Dr. C. E. Myers of the State Board of Education in Virginia; Supervisor R. V. Bagby of Roanoke College, Virginia; Dr. James F. Bursch of Sacramento, California; Dr. Virgil Smith of Seattle, Washington; Dr. C. W. Odell of the University of Illinois and the Superintendent of the schools of Urbana, Illinois; and to Dr. E. S. Henry of the New York University, New York City.

²The printing was done through the courtesy of Ohio State University and without expense to the Investigation.

TESTS AND MEASUREMENTS

TABLE I: Number of Cases for Each Grade on Each Form of the Preliminary Tests, and the Geographical Distribution of the Cooperating Cities and Towns.¹

	Form A					Form B				
	Grades					Grades				
	4	6	8	10	12	4	6	8	10	12
Beverly, Mass.....	163	143	191	164
Charlestown, W. Va....	...	97	116	156	136	...	94	115	123	134
Columbus, Ohio.....	120	126
Flagstaff, Ariz.....	...	11	10
Grosse Pointe, Mich...	271	...	149
Hamtramck, Mich.....	...	196	824	340	176
Highland Park, Mich..	150	244	197	189
Lewiston, Idaho.....	...	93	124	90	63	...	94	127	90	60
Lincoln, Neb.....	177	157	299
Mechanicville, N. Y...	135	150	108	48
Milwaukee, Wis.....	326	...	70
Montclair, N. J.....	...	22	37	...	23	22
New York City.....	46	39
Penn Yan, N. Y.....	100	72	34	60
Sacramento, Cal.....	713	720
So. Hamilton, Mass...	32	25	22	27	...	30	25	24	26	...
Wenham, Mass.....	14	13	10	14	12	12
	212	457	2416	926	1018	209	483	1397	926	976

	Form C					Form D				
	Grades					Grades				
	4	6	8	10	12	4	6	8	10	12
Americus, Kans.....	...	11	7	23	15	...	11	7	22	15
Atlanta, Ga.....	266	270
Berkeley, Cal.....	191	182
Beverly, Mass.....	141	184	138	165
Columbus, Ohio.....	128	117
Emporia, Kans.....	...	32	64	45	26	...	32	65	22	14
Houston, Tex.....	98	99
Kansas City, Mo.....	205	247
Long Beach, Cal.....	...	254	265	228	248	...	250	230	230	247
New York City.....	43	41
Philadelphia, Pa.....	152	91	146	88
Roanoke, Va.....	473	464
Seattle, Wash.....	...	125	146	122	143
So. Hamilton, Mass...	29	26	25	23	...	28	26	24	25	...
Urbana, Ill.....	276	274
Wenham, Mass.....	14	12	12	15	14	11
	214	460	1534	817	1100	201	455	1488	583	810

¹In addition were received, too late to be included in the tables, 468 cases, eighth grade, Preliminary Form A and 470 cases, eighth grade, Preliminary Form B from Denver, Colorado, and 27 cases, tenth grade, Preliminary Form D, from Cambridge, Massachusetts.

METHODS OF PROCEDURE IN THE INVESTIGATION

each child actually marked was a 3 by 5 inch card¹ on which there appeared spaces for writing down the letter that stood before the answer he selected as right. On this card he wrote also his name, age, grade, and other such information. A copy of the first two lines of this card appears below:

NAME					46					47					48					49					50									
CLASS					51					52					53					54					55									
DATE					56					57					58					59					60									
TEST					61					62					63					64					65									
SCORE					66					67					68					69					70									
1					2					3					4					5					6					7				
6					7					8					9					10					11					12				
11					12					13					14					15					16					17				
16					17					18					19					20					21					22				
21					22					23					24					25					26					27				
26					27					28					29					30					31					32				
31					32					33					34					35					36					37				
36					37					38					39					40					41					42				
41					42					43					44					45					46					47				
46					47					48					49					50					51					52				
51					52					53					54					55					56					57				
56					57					58					59					60					61					62				
61					62					63					64					65					66					67				
66					67					68					69					70					71					72				
71					72					73					74					75					76					77				
76					77					78					79					80					81					82				
81					82					83					84					85					86					87				
86					87					88					89					90					91					92				
91					92					93					94					95					96					97				
96					97					98					99					100														

PRESSEY POCKET TESTER

© AMERICAN EDUCATION PRESS, Inc., Columbus, Ohio

The value of this technique is twofold. The cards are much less expensive than a separate blank for each child. And the labor of scoring and tabulating is greatly reduced. Thus the entire 16,682 cards² were scored and tabulated by item for a total cost of \$255.65, or barely 1½ cents a card (.0153¢). In contrast it may be said that the same scorers similarly handled results from the

¹Printed by the American Education Press, Columbus, Ohio.

²There were 3,615 other cards used in odd-numbered grades; these were scored but not tabulated by item. They were not included in the above number.

TESTS AND MEASUREMENTS

TABLE II: Distribution of Total Scores on Preliminary Form A of the Vocabulary Test

Score	GRADES				
	4	6	8	10	12
80.....
78-79.....	5	37
76-77.....	5	16	91
74-75.....	5	18	123
72-73.....	15	42	156
70-71.....	27	58	123
68-69.....	35	79	89
66-67.....	66	64	98
64-65.....	..	1	93	57	75
62-63.....	93	82	50
60-61.....	..	2	128	71	51
58-59.....	1	1	132	69	30
56-57.....	..	5	141	55	25
54-55.....	..	3	130	48	15
52-53.....	..	5	137	46	15
50-51.....	..	8	115	50	5
48-49.....	1	7	106	52	8
46-47.....	..	16	119	27	7
44-45.....	2	10	131	21	6
42-43.....	..	18	120	12	2
40-41.....	1	20	111	11	..
38-39.....	..	19	116	10	2
36-37.....	1	22	105	11	2
34-35.....	..	18	94	12	..
32-33.....	2	30	77	3	3
30-31.....	5	44	96	1	1
28-29.....	4	27	72	2	1
26-27.....	5	39	47
24-25.....	8	41	47	1	1
22-23.....	13	31	21
20-21.....	10	28	13	1	2
18-19.....	22	12	9	1	..
16-17.....	32	14	5	1	..
14-15.....	22	16
12-13.....	18	9	1
10-11.....	13	2	2
8-9.....	16	3
6-7.....	11	4
4-5.....	11	1
2-3.....	6	1
0-1.....	8
Total.....	212	457	2,416	926	1,018
Median.....	16.0	30.0	48.4	60.8	70.4

RESULTS OF THE INVESTIGATION

two school systems in which the blanks were marked at a rate of $4\frac{1}{2}$ cents per test; this rate would have required \$750.69 for handling the same total number of cases (plus the greater expense in printing, since each child would have needed a blank).

IV. RESULTS OF THE INVESTIGATION

The record cards were first scored for the total number of items correct. The results from a single form¹ are presented on p. 164, to show the nature of these results. The wide variation of scores should be especially noted as indicative of the individual differences in attainment shown among children classified by the school as being of the same educational level. Thus, in the fourth grade there is one child who tests near the median for the tenth grade, another who tests at the median for the eighth, and eleven who test above the median for the sixth. In the tenth and twelfth grades together there is one child who tests at the median for the fourth grade, an additional nine who test at or below the median for the sixth, and one hundred and thirty-one more who test below the median for the eighth. In Grade VIII the range is from almost complete failure to almost complete success. The medians also are interesting. The largest gain is between the sixth and eighth grades, with smaller gains between the fourth and sixth, the eighth and tenth, and the tenth and twelfth. This characteristic of the results will be returned to later.

While the separate forms of the tests have no significance because of the largely chance arrangement of the items, the sum of the medians on the four forms will give a rough indication of the degree of mastery of the 310² items in the different grades. Since with a very few exceptions, the same children did not take all tests, no total score on the entire examination is possible, but an approximation may be arrived at by adding together the medians for the four forms; this total should not be far from the truth, since

¹Distribution for all forms will be found in Appendix C at the end of the paper.

²Of the 325 original items, 10 were used as examples, 3 were unsatisfactory, and 2 had to be dropped because of an error in the score sheet.

TESTS AND MEASUREMENTS

the results for all forms are based on equally unselected groups. The results of this addition appear below:

TABLE III: Summation of the Medians for Each Grade on Each Preliminary Form of the Tests¹

	Grade 4	Grade 6	Grade 8	Grade 10	Grade 12
Form A.....	16.0	30.0	48.4	60.8	70.4
Form B.....	26.1	39.2	55.5	63.4	70.3
Form C.....	18.3	32.0	45.9	57.2	68.2
Form D.....	22.0	37.4	54.0	61.8	71.4
Total.....	82.4	138.6	203.8	243.2	280.3
Difference.....		56.2	65.2	39.4	37.1

¹It will be noticed that the forms are not altogether equal in difficulty. However, the largest variation is due to the fact that Form C had only 74 items that were scored, and Forms A and D 79 items, while Form B had 80. Any further variation is due to the fact already mentioned that the assistant in making up the forms did not use an altogether random selection. Since the total score on any one form is of no special significance the matter is of little importance, but it does explain the variability in the medians.

In considering the increase from one of these levels to the next it is clear that the main improvement takes place between Grades IV and VIII, the additional amounts being only a little more than the rise from Grade VI to Grade VIII by itself. This failure of the high school grades to show more marked gain is not due to a lack of room for improvement; the Grade XII median is still thirty points below the total number of items in the examination. The lack of growth in vocabulary in the upper grades is especially significant because most pupils have one year of history in high school and many of them have two (plus more or less work in other social sciences). The situation is probably due in part to a difference in attitude toward the problem of vocabulary; the teachers in the elementary grades assume few words as known and consequently make specific efforts to teach vocabulary, whereas those in high school tend to ignore the problem. It is also natural that, the easiest and commonest words having been learned, mastery of the harder and less frequent words should be slow.

After the record cards had been scored and the total scores handled as shown above, the correct answers were tabulated by item for each grade in each school. The number of children pass-

RESULTS OF THE INVESTIGATION

ing each item in each grade on each form for all schools was then found and these numbers turned into percentages of children in each grade knowing each item. At this point the items from the various forms were rearranged and the responses classified in terms of the words being tested. The list of 346 words, knowledge of which was required in the 310 items,¹ with the per cent recognizing each word in each grade began as follows:

TABLE IV: The Percentage of Children in Each Grade Recognizing the Meaning of Each Word

Words	Grade 4	Grade 6	Grade 8	Grade 10	Grade 12
ambassador.....	(used as example and not scored)				
authorities.....	21	40	81	87	94
consul.....	19	38	61	83	87
governor.....	61	61	84	88	97
king.....	(item omitted because of error in scoring key)				
minister.....	11	26	32	32	51
etc.					

It should be noted that this list of words corresponds exactly to the list previously referred to as the outcome of the research into what constitutes the essential terms, which list is presented in full at the end of this paper. Each word in that list is accounted for in some way in this list of results, which appears in its entirety later on, p. 208.

The next step was to consider once again, from a study of the gain in percentages shown by the 346 words tested, the general growth in vocabulary. The percentages shown by all words in each grade were therefore tabulated, with results as shown in Table V.

This table shows the same general trend as the earlier ones—that the most significant rises come in the lower grades. It shows also the spread of vocabulary within a given grade. For instance, there are three words known by 80 to 84 per cent of the pupils even as low as the fourth grade, and another twenty-one known by over 50 per cent. On the other hand, ten words were recognized by less

¹It should be pointed out that when two terms are tested by the same item the resulting percentages reflect presumably the mastery of the harder word.

TESTS AND MEASUREMENTS

TABLE V: Percentage of Children in Each Grade Answering Each Item Correctly; also the Spread of Mastery of the 346 Words in Each Grade

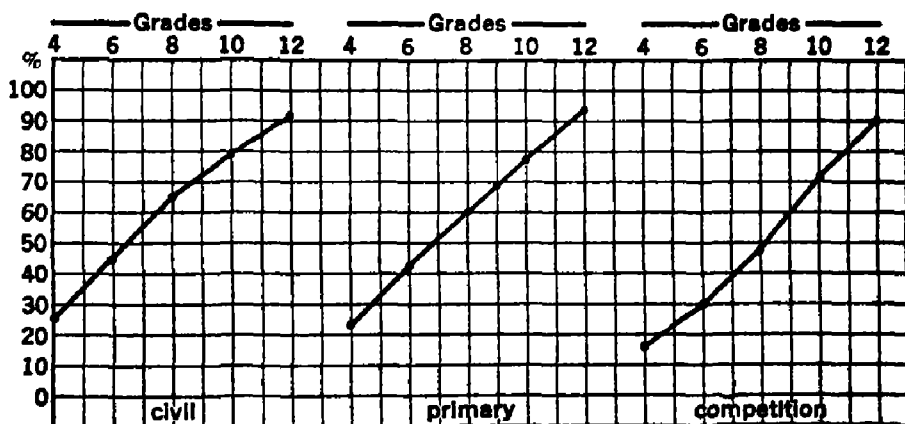
Per cent	GRADES				
	4	6	8	10	12
100.....	1
95-99.....	..	2	22	43	110
90-94.....	..	3	33	63	61
85-89.....	..	8	29	51	45
80-84.....	3	17	36	24	33
75-79.....	2	9	26	27	27
70-74.....	1	22	24	24	15
65-69.....	1	11	15	21	12
60-64.....	5	24	22	13	11
55-59.....	3	12	23	24	8
50-54.....	9	23	21	14	7
45-49.....	11	23	18	7	4
40-44.....	19	29	22	8	2
35-39.....	23	26	11	9	5
30-34.....	36	31	8	12	..
25-29.....	48	36	15	1	1
20-24.....	41	30	9	3	4
15-19.....	56	21	9	1	..
10-14.....	51	15	3	1	..
5-9.....	27	4
0-4.....	10
Total Number of Items.....	346	346	346	346	346
Percentage of Chil- dren Knowing Median Item....	23.5	41.7	69.0	81.6	90.0

than 5 per cent of the pupils. The sixth grade shows words varying from two known by 95 to 99 per cent of the children to four known by less than 10 per cent. It is interesting to note that even in the tenth and twelfth grades there are still words known by less than half of the pupils—42 such words in Grade X and 16 in Grade XII. Although the percentage of children knowing the median word for each grade rises definitely, there is a marked lagging behind of certain words. If these terms could be isolated and taught, the mastery could be considerably improved with relatively little effort. At the other extreme, it is encouraging to see that by the end of high school 172 words are known to 90 per cent or more of the children.

RESULTS OF THE INVESTIGATION

If one were to consider as really mastered those words recognized by 90 per cent or more of a class, then there would be no such terms in Grade IV, five in Grade VI, fifty-five in Grade VIII, one hundred six in Grade X, and one hundred seventy-two in Grade XII. With this standard¹ almost exactly half the total 346 words were "really known" at the end of high school.²

As a final step in presenting the results the writer constructed graphs for each word, using the percentage of children passing each item in each grade as a basis. The entire number of graphs will not be presented because they would occupy too much space, because they are of a limited number of types easily demonstrated by a selection, and because anyone interested in studying them can easily construct them for himself from the percentage table given at the end of the paper—the same table the author used. However, certain typical graphs are given below, so that the general nature of the results may be appreciated. It should be noted that these graphs show learning curves for the term being tested. If the reader wishes to know the exact context of the terms thus illustrated, he is referred to the table of percentages in Appendix D, where he will find the designated item in which each word is to be found; he can thus look up the actual questions on the test blanks, which also appear at the end of the paper, and have the question before him.



¹Anyone wishing to set a standard at a point other than 90 per cent can do so by use of the detailed table giving percentages in every grade for every word.

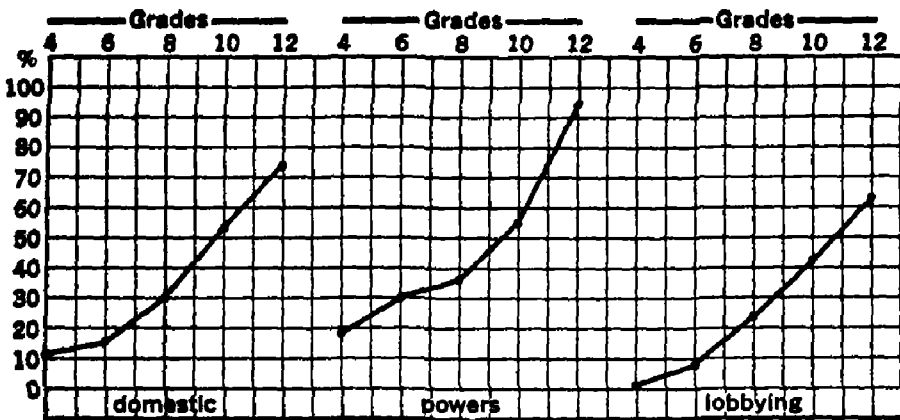
²A list of these words appears later, p. 175 ff.

TESTS AND MEASUREMENTS

The first group of graphs illustrates a common type of learning curve—the straight line growth from low mastery in Grade IV to high mastery in Grade XII. There were seventy such graphs.

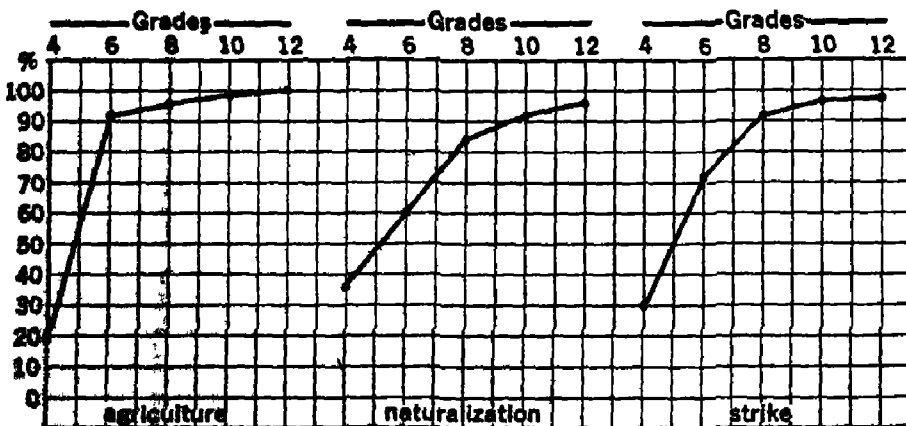
These three terms show gains from the bottom to the top of 65, 71, and 74 respectively.

The next four sets of graphs show variations on the above fundamental type. The graphs for the learning of the words “domestic,”



“powers,” and “lobbying,” show a slow start, but a straight line increase once the learning has begun. These concave curves (in number 31) are typical of the more difficult words.

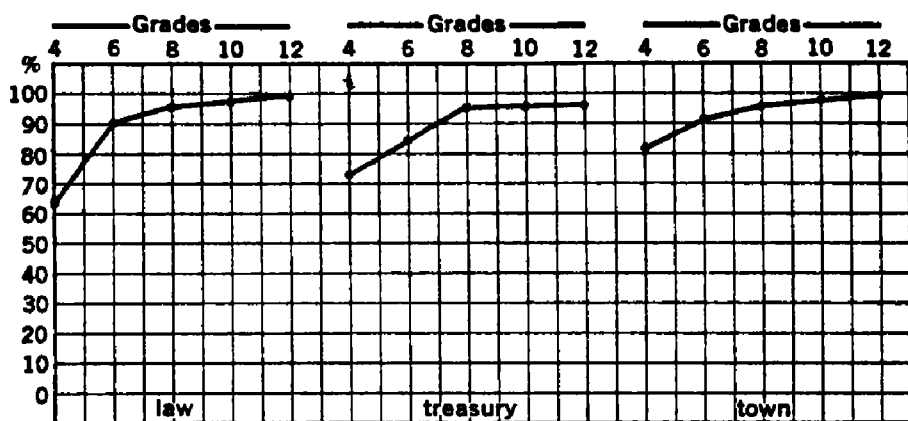
A variation in the opposite direction is shown by the curves representing mastery of the words “agriculture,” “naturalization,” and “strike.” For these words the growth comes in the early



RESULTS OF THE INVESTIGATION

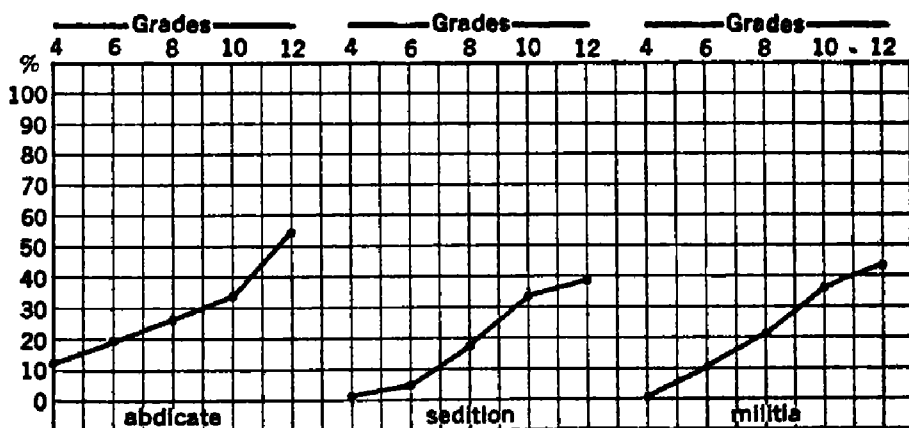
grades with almost perfect recognition thereafter—thus giving a convex curve of which there were 32 instances.

A third variation is illustrated by the three curves below.



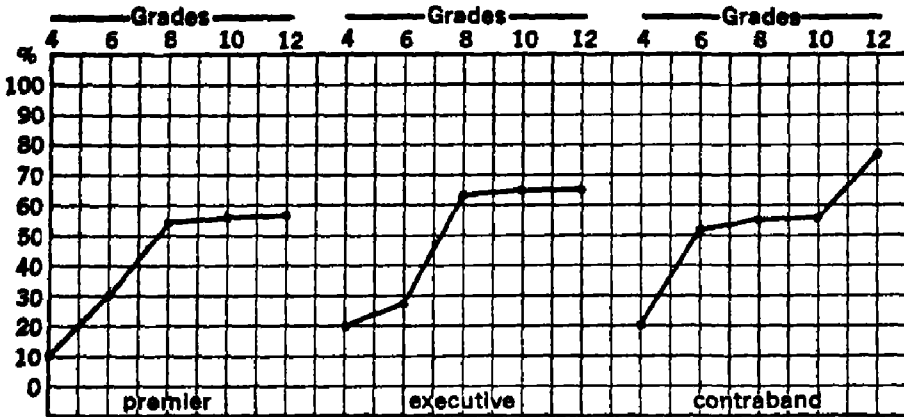
The mastery of these three words is already so high in Grade IV that little growth is possible, but what growth there is takes place regularly. They represent really the top of a learning curve whose beginnings are somewhere below Grade IV. Words showing a mastery of 40 per cent or more in the fourth grade and then increasing regularly are included in this group; there were 45 such words.

The words "sedition," "abdicate," and "militia" illustrate one more variation. In these cases the growth is again perfectly regular, but so slow that the final achievement in no instance exceeded 54 per cent. There were 36 curves of this type.



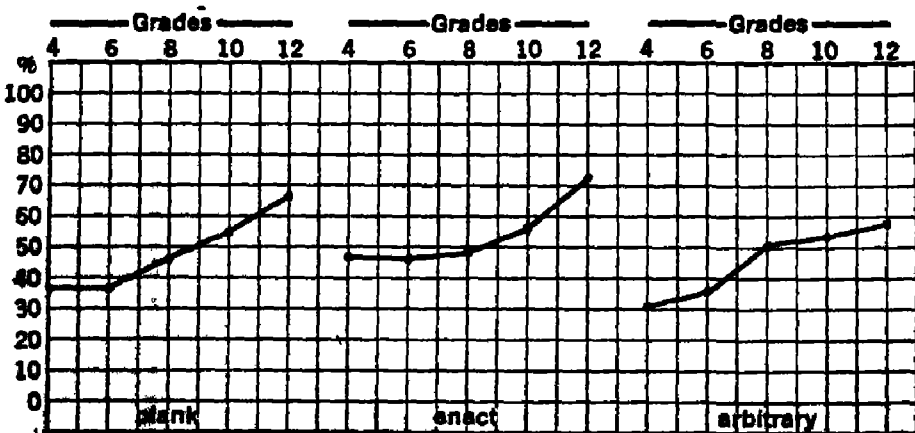
TESTS AND MEASUREMENTS

The curves thus far have shown regular growth. There remain four more types of curves in which growth is, for some reason, more or less irregular. Those for the words "premier" and "executive" contain marked plateaux. What learning took place oc-



curred early, the later grades merely holding their own. In the case of "contraband" there is again a plateau, but this time from the sixth through the tenth grade, with Grade XII marking a final slight rise. There were 19 curves showing plateaux at some point in their rise (and below 90 per cent, after which a plateau is to be expected).

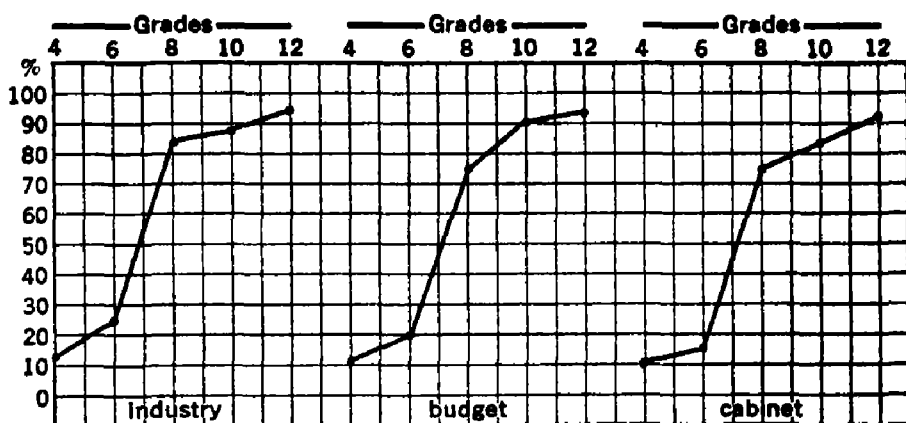
The three curves next presented are interesting because of the failure of the upper grades to make the expected gain. The Grade IV results start these curves off with a relatively high achievement



RESULTS OF THE INVESTIGATION

—from 31 per cent to 47 per cent in this earliest grade knowing these words. In spite of this good start, the final standing in no case exceeds 73 per cent. The total gains are only 26, 30, and 27 per cents. These words need study to find out why they do not show more rise. There were only 6 words that acted in this way.

The next group of curves (which contains 80 instances) illustrates a rate of learning commonly found in this study wherein the rise in total score from the sixth to the eighth grade always exceeds that between any other two grades. There are more curves of this shape than of any other. In each of the three instances

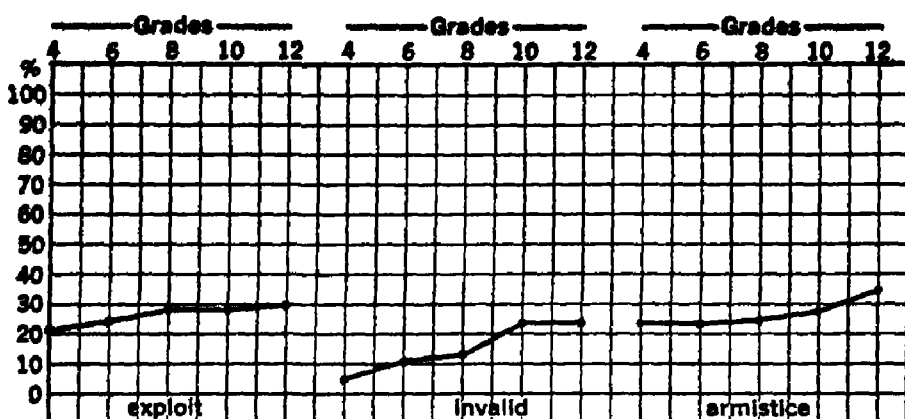


there is little learning before Grade VI and not a great deal after Grade VIII, but an increase of 50 per cent or more between these two grades. If the reasons for the splendid achievement at this level could be isolated, and the techniques applied to learning at the higher levels the problem of vocabulary might be solved.

Finally, there are few words showing such results as are illustrated below. These can hardly be called learning curves in the absence of both learning and curve. The mastery in all grades is low, nowhere exceeding 35 per cent. As may be seen from the table already presented (Table V) these results are not numerous—in fact, there were only 14 curves of this type.

In addition to these well-defined types of curves, there are a few words that show peculiar results in that a later grade scores lower than previous grades. There seems no plausible explanation

TESTS AND MEASUREMENTS



for these findings. There are, however, only 13 such words, and in no case is there more than one grade that shows such variation. All other words show every grade at least holding its own with the grade below and usually showing improvement.¹

Consideration of any or all of the results presented above leaves one with the conviction that not enough special vocabulary is acquired by the average student at any level to serve as a sure foundation for the reading of the average textbooks used in these grades. Again it should be pointed out that these words are, for the most part, of fundamental character. Words less frequently used and less important would presumably show even less mastery. The achievement shown is, then, the "cream"—and one wonders what the skimmed milk would be like.

V. PRACTICAL USEFULNESS OF THE DATA

In order to make the results of practical use the writer has constructed the following lists, which contain those words known to at least 90 per cent of the pupils in each of the grades tested. A textbook writer would probably be safe in using these words without explanation. Naturally, he would use many others, partly because they would be necessary to express his ideas and partly because only through the reading of new words will a child's vocabulary be increased. The writer would like to suggest, however,

.. ¹A classification of the curves appears at the end of Appendix D.

PRACTICAL USEFULNESS OF THE DATA

that words showing a lower degree of mastery than those listed should be introduced with some kind of explanation as to their meaning. Similarly, teachers could use the lists in their teaching, being especially careful to explain the words that do not occur on the lists below. In general, not only this list but the detailed list showing percentages may be used as guides to what can be expected of children at different levels.

TABLE VI: Words Known by Ninety Per Cent or More of the Pupils in Each Grade Tested

Grade 6 ¹	Grade 8	Grade 10	Grade 12
agriculture	agriculture	agriculture	abolitionist
law	bankrupt	alien	adjourn
submarine	boundary	alliance	administration
town	century	amendment	agriculture
voyage	coinage	anti-slavery	alien
	committee	aristocrat	alliance
	communication	assessment	allies
	community		ambassador
	conference	ballot	amendment
	congress	bankrupt	annex
	continent	bond	anti-slavery
	court	boundary	appoint
		budget	aristocrat
	defensive		article
	discover	candidate	assembly
		capitol	assessment
	expedition	census	authorize
	export	century	
		coinage	ballot
	fleet	colony	bankrupt
	freedom	commerce	bond
		committee	boundary
	goods	communication	bribery
		community	budget
	House of Representatives	conference	
		congress	cabinet
	import	continent	campaign
	irrigation	court	candidate
		credit	capitol
	law	currency	census
	liberty		century
		decision	civil
		defensive	coinage
	majority	democrat	colony

¹ There were no such words in Grade 4.

TESTS AND MEASUREMENTS

TABLE VI: Words Known by Ninety Per Cent or More of the Pupils in Each Grade Tested—*Continued*

Grade 6 ¹	Grade 8	Grade 10	Grade 12
	manufacture merchandise missionary mob	discover document emigration empire employee expedition export	commerce committee communication community competition compromise conciliate conference congress constitution continent council court credit
	nationality naval offensive patriotism peasant pirate political party population	finance fleet freedom goods	
	senate settlement slave stocks strike submarine surrender	greenback House of Repre- sentatives immigration import inaugurate inhabitants invention investment irrigation	creed crime crusade currency current
	tax territory town transportation treasury treaty	labor law liberty	decision defensive delegate democracy democrat discover doctrine document draft duty
	veteran vice-president violation volunteer voyage	majority manufacture market massacre merchandise missionary mob nationality native naturalization natural resources naval navigation neutrality	election emigration empire employee enemy execution exile expedition export finance fleet foreign fraud freedom

¹There were no such words in Grade 4.

PRACTICAL USEFULNESS OF THE DATA

TABLE VI: Words Known by Ninety Per Cent or More of the Pupils in Each Grade Tested—*Continued*

Grade 6 ¹	Grade 8	Grade 10	Grade 12
		offensive official opponent	goods governor greenback
		parliament patriotism peasant pirate polls population prohibition prosperity public opinion	hostile House of Represen- tatives
		republican	illegal immigration impeach import inaugurate industry inhabitants internal invention investment irrigation
		senate session settlement slave standard of living stocks strike submarine surrender	jury justice
		tax territory town trade transportation treason treasury treaty	labor law legislative liberty
		unanimous union	majority manufacture market massacre merchandise migration military mint missionary mob
		verdict veteran vice-president violation volunteer voyage	nation nationality native naturalization natural resources naval navigation neutrality

¹There were no such words in Grade 4.

TESTS AND MEASUREMENTS

TABLE VI: Words Known by Ninety Per Cent or More of the Pupils in Each Grade Tested—Continued

Grade 6 ¹	Grade 8	Grade 10	Grade 12
			nullify
			offensive
			official
			opponent
			parliament
			patriotism
			peasant
			pioneer
			pirate
			political party
			polls
			population
			primary
			proclamation
			prohibition
			proposal
			prosperity
			protective
			province
			public opinion
			raw material
			reparations
			republic
			republican
			senate
			session
			settlement
			slave
			smuggle
			standard of living
			state
			stocks
			strategic
			strike
			submarine
			suffrage
			supreme court
			surrender
			tax
			territory
			testimony
			town

¹There were no such words in Grade 4.

PRACTICAL USEFULNESS OF THE DATA

TABLE VI: Words Known by Ninety Per Cent or More of the Pupils in Each Grade Tested—*Continued*

Grade 6 ¹	Grade 8	Grade 10	Grade 12
			trade
			transportation
			treason
			treasury
			treaty
			unanimous
			union
			verdict
			veteran
			veto
			vice-president
			violation
			volunteer
			voyage
			witness

¹There were no such words in Grade 4.

There are in existence a number of frequency counts of the special vocabulary in various textbooks for use in elementary school, high school, or college. Anyone interested in so doing can check one or more of these counts—which show the actual number of appearances of each technical word in a particular book—against the percentage knowing each term, or against the level at which each term is really learned. The writer will present one such analysis and compare it with the results, for the purpose of showing how this practical use of the data might be accomplished. The words occurring in the highest fifth of the frequency distribution² in a textbook widely used in Grades XI and XII were selected for study. Although this book is recommended chiefly for the final review of American History in high school, it is occasionally used as

²These words were in the highest fifth of the distribution table, but are not one-fifth of the total number of words. They were selected by taking the entire variation of frequency shown on the table—from over 300 to 1—dividing this distance by 5, and then copying down those words that had a frequency from 300 to 240—the highest fifth. Over half the words had a total frequency of either 1 or 2; the words selected constitute only 16 per cent of the total number of words—which was 862.

TESTS AND MEASUREMENTS

low as the eighth grade, and with fair frequency in Grades IX or X, in case the subject is taught in those grades. There were 137 words selected as above described. Of these 137, 22 were not included in the tests (one would not expect perfect agreement between the words in a list based merely on frequency in a single book and words based on analysis of several books, plus newspaper counts, plus judgment of teachers), 3 were used as examples, and 2 appeared in unsatisfactory items, thus leaving a total of 110 words for which results were obtained on the tests. The percentage of children in Grade XII recognizing these 110 words is shown in the table below.

TABLE VII: Percentage of Twelfth Grade Pupils Recognizing Each of 110 Words that Occurred in the Highest Fifth of the Frequency Distribution for a Single Textbook

Per Cents	Number of Words Known
100.....	..
95-99.....	48
90-94.....	24
85-89.....	12
80-84.....	11
75-79.....	5
70-74.....	2
65-69.....	1
60-64.....	3
55-59.....	1
50-54.....	3
45-49.....	..
40-44.....	..
35-39.....	..
30-34.....	..
25-29.....	..
20-24.....	..
15-19.....	..
10-14.....	..
5-9.....	..
0-4.....	..
Total.....	110

Seventy-two words (or 66 per cent) were known by 90 per cent or more of the children, but 38 (or 34 per cent) were known by less than that proportion. Ten words (or 9 per cent) were recognized by less than three-fourths, but more than one-half, of these high school

PRACTICAL USEFULNESS OF THE DATA

seniors. There was a total of 862 words found in this text; it is reasonable to suppose that the mastery of the remaining 725, which had frequencies somewhere below the upper fifth of the distribution table, would be appreciably less. It would seem, then, that at the end of high school these pupils were not entirely ready to read texts using even the most commonly appearing words. If this text were used in Grade VIII, there would be only 22 of the 110 words studied known to 90 per cent or more of the children; for Grade X there would be 53 such words. Although both frequency counts and objective tests have their limitations as educational techniques, this comparison between them seems to leave no doubt as to the lack of foundation in the matter of vocabulary upon which this textbook rests—even at the grade for which it was intended.

A similar study was next made showing twelfth grade mastery of the 133 words found in the highest fifth of the frequency distribution in the author's own counts of special vocabulary in newspapers

TABLE VIII: Percentage of Twelfth Grade Pupils Recognizing Each of 101 Words that Occurred in the Highest Fifth of the Frequency Distribution of Words Found in Newspapers and Current Magazines

Per Cent	Number of Words
100.....	..
95-99.....	41
90-94.....	20
85-89.....	9
80-84.....	10
75-79.....	5
70-74.....	3
65-69.....	2
60-64.....	5
55-59.....	3
50-54.....	3
45-49.....	..
40-44.....	..
35-39.....	..
30-34.....	..
25-29.....	..
20-24.....	..
15-19.....	..
10-14.....	..
5- 9.....	..
0- 4.....	..
Total.....	101

TESTS AND MEASUREMENTS

and current magazines. Of these 133 words, 26 did not appear in the test (most of them being too easy for such inclusion), 5 were used as examples, and 1 appeared in an unsatisfactory item. The remaining 101 showed degrees of understanding by twelfth grade pupils as shown on Table VIII.

Of these 101 terms for which test results were obtained, 61 (or 60 per cent) were recognized by 90 per cent or more of the pupils; 24 were known by some proportion, between 90 and 75 per cent; the remaining 16 were known by less than 75, but more than 50, per cent. This showing is not quite so good as that for the textbook results just presented. It suggests that the high school graduates are not adequately prepared to read the current materials through which they can most readily follow contemporary history. There was a total of 1,010 other words found with lower frequencies in the same newspapers; these would presumably show less adequate mastery. There seems no escaping from the fact that the vocabulary problem is far from being solved and that preparation of assignments in history or reading of the newspapers is likely to continue in *status quo* until more meanings are acquired by the pupils.

It is the writer's hope that others will find the data of value in other and equally interesting comparisons. The sole object in collecting the detailed information presented in this paper is to have it *used*.

VI. CONCLUSIONS AND CONSTRUCTIVE SUGGESTIONS

It remains to point out certain relevant lines along which further work needs to be done if the problem of vocabulary is to be adequately solved.

(1) There needs to be further study of the words revealed by word counts so as to get them classified in such a way as to be directly useful to teachers. The writer would suggest a division of all the special words found into "essential" and "accessory" lists (on the joint basis of frequency, importance, and sociological value) and a further classification into levels of difficulty on the basis of these and further test results (plus, preferably, a logical

CONCLUSIONS AND CONSTRUCTIVE SUGGESTIONS

grouping within each level). Such work would eventuate in lists for each grade, or at least each level in the school system, so that the teacher could readily see (a) what words pupils at a given stage of development can reasonably be expected to know and (b) of these words which are the more fundamental, and should be taught first. Such lists would have a practicability (and would therefore get into actual usage in the classroom) that could never be achieved by a long, alphabetical list—which would, in all probability, be used almost exclusively for research purposes.

(2) Much further study needs to be made of the validity—not the reliability, which is known to be high—of objective test items. Most people seem either to assume that words correctly recognized in such a test will automatically be known if met in another context, or else take an equally extreme position and state that the objective test tells nothing whatever about recognition in any other situation. Presumably neither group is right—and in any case both must be merely rationalizing a point of view since nobody knows the truth. It is the writer's guess that the objective test reflects actual knowledge that is transferred to reading situations more often than not; but only extensive and impartial investigation can settle this question.

(3) Since school children obviously need to be taught the special vocabulary of history by some more adequate means than are now employed, the writer would like to suggest that texts should be worked out—certainly for grades below the college level—with careful study as to the introduction of new words. Thus it might be possible to introduce one new word on every other page, or something of the sort, with careful explanation of its meaning. In connection with the word lists above mentioned, the texts might become media for special training in vocabulary as well as for transmission of historical information. A simply-worded glossary at the back of texts, with footnote references for difficult or new words, would be a most welcome addition. In any event, some technique needs to be worked out so that texts will not dump a mass of unexplained—and hence largely unassimilated—words upon pupils.

TESTS AND MEASUREMENTS

(4) A consideration of the frequency counts from almost any text in history leaves one with the feeling that there is an enormous and unnecessary vocabulary burden in the form of words used only once or at most twice in the entire book. Such words are usually skipped by the pupil and, of course, such infrequent use does not permit of any real mastery. Nor can pupils be expected in one year to look up and learn the meanings of a considerable proportion of 862 words—to use the text mentioned in the previous section—and at the same time acquire a year's worth of history. Much of the dispersion of vocabulary in current texts seems to be due to efforts at style by the persons who write the textbooks. The writer would venture a guess that the style is largely wasted on the public school child and serves chiefly to add elements of confusion in the form of almost literally hundreds of special and difficult words used too infrequently to permit learning and with such profusion as to discourage all but the most persistent. If the vocabulary could be concentrated by using fewer words and using them oftener, there would be sufficient practice given to promote real mastery of this more restricted, but more fundamental, vocabulary. Naturally, such concentration would be correlated with lists of essentials to be acquired at the successive levels. Such a concentration of vocabulary has already been tried out with success in the construction of reading materials in the elementary grades for the mastery of general reading vocabulary. Each word of the most commonly used thousand (or two thousand or three thousand, according to the grade for which the book is being written) is repeated in the fundamental readers for a given grade in proportion to both its frequency of appearance in general reading matter and its difficulty. Learning to read becomes, with materials of this sort, a controlled process rather than one that proceeds on the basis of mere hope and courage. If history texts for the public schools could be rewritten upon such a basis, learning to read them would be not only easier—which is, perhaps, beside the point—but infinitely more successful.

(5) Naturally, one does not wish children to acquire words for the sake merely of knowing more words, but in order that they

CONCLUSIONS AND CONSTRUCTIVE SUGGESTIONS

may read and think more efficiently. The writer has encountered a number of people who are inclined to scoff at any study of vocabulary on the basis that such study stresses "mere words." Quite true. But no one can either read or think without them. Word meanings are the raw materials of constructive thinking, and no pupil is going to think accurately without an adequate vocabulary. Anyone doubting the importance of "mere words" in the preparation of assignments in history should sit down beside a probation student, watch him as he fumbles around with his reading, and investigate his thinking. After one finds out that he thinks "anarchy" was the outstanding form of government in ancient Greece, "reciprocity" is a form of revenue, and a "budget" is a statement of party principles, one begins to get a new sense of the necessity of mastering elementary meanings before advancing to higher realms of thought.

Further light is shed on the essential need for adequate word knowledge if one will have patience with the stumbling probationer, will teach him the words he should have learned long before, and will then watch him gain markedly in his thinking—now that he has something with which to think. The writer would like to stress in closing, the strategic importance of the vocabulary problem, especially in the social sciences, because the words there learned are directly transferable to any reading of current events, and to urge that constructive steps be taken to bring about a learning that will act as an adequate basis for the development of far better reading and thinking than could be done with a vocabulary no more adequate than that revealed as typical by the results of the present investigation.¹

¹The writer has just completed the preparation of two final vocabulary test forms of 140 scored items each, covering the essential terms. Since the entire learning curve for each question was available, the forms have been constructed by pairing the items according to the shape of the entire curve. Thus, two items showing percentages passed in the grades as follows:

	Grade 4	Grade 6	Grade 8	Grade 10	Grade 12
Item (1).....	14	21	49	58	72
Item (2).....	17	23	50	54	75

TESTS AND MEASUREMENTS

APPENDIX A: Words Selected as the Essential Core of Vocabulary in History

A. Governmental Terms:

ambassador	democracy	bill	appropriation
authorities	empire	declaration	appointment
consul	federal	decree	budget
governor	government	document	*coinage
king	*imperialism	law	currency
minister	monarchy	legislation	customs
official	republic	#measure	debt
#prime-minister	self-government	petition	duty
premier	tyranny	proclamation	expenditures
#police	union	proposal	*greenback
president		#provision	*mint
representative	city	#report	protective
secretary	colony	resolution	revenue
senator	#country	restriction	tariff
sovereign	county	statute	tax
#statesman	*dominion		treasury
vice-president	nation	#abolish	
	province	*abdicate	doctrine
assembly	state	adjourn	issue
bureau	territory	annex	policy
#board	town	appoint	#reservation
cabinet		authorize	
commission	alliance	compromise	centralization
committee	arbitration	concede	civil
conference	diplomacy	*conciliate	civil service
Congress	foreign	*confiscate	domestic
Council	international	enact	internal
department	negotiation	enforce	interstate
House of Rep-	neutrality	#grant	local
resentatives	#pact	impeach	municipal
league	peace	inaugurate	states rights
legislature	powers	nullify	
parliament	reciprocity	ratify	administration
senate	treaty	repeal	regime
session		*repudiate	
	amendment	*sanction	*capitol
anarchy	article	veto	patriotism
*commonwealth	charter		
#communism	constitution	executive	prohibition
confederacy	act	legislative	reconstruction
despotism		judiciary	referendum

B. Political Terms:

campaign	*anti-slavery	majority	ballot
candidate	*abolitionist	#minority	election
caucus	democrat	unanimous	polls

*This symbol is used to indicate those words that the author would now, on the basis of further research, omit from a list of essentials.

#This symbol is used to indicate those words that were, for one reason or another, not included in the test, although their essential character is clear.

CONCLUSIONS AND CONSTRUCTIVE SUGGESTIONS

APPENDIX A: Words Selected as the Essential Core of Vocabulary in History—Continued

#convention	*federalist	conservative	primary
deadlock	political party	*partisan	suffrage
delegate	progressive	radical	vote
nominate	republican		
opponent	socialist		
plank	*whig	lobbying	
platform		patronage	
politics		spoils system	
ticket			
C. Economic Terms:			
business	manufacture	employee	inflation
commerce	merchandise	#employer	investment
commodity	production	labor	market
company	#property	strike	panic
competition	raw material	union	speculation
consumer	rebate		stocks
exploit	#shipping	#bank	
export	trade	bankrupt	communication
#factory		bond	public utilities
goods	corporation	capital	transportation
import	monopoly	credit	
industry	trust	crisis	prosperity
		depreciation	wealth
		finance	
D. Sociological Terms:			
aristocrat	rural	emigration	#education
peasant	urban	expansion	#institution
slave		immigration	invention
#society	census	migration	reform
	inhabitants		
community	population	emancipation	#people
homestead		freedom	#private
pioneer	negro	independence	#public
plantation	race	liberty	public opinion
settlement		oppression	standard of living
	mob		
	riot		
E. Legal Terms:			
*arbitrary	nationality		fraud
illegal	native	jury	graft
*invalid	naturalization	testimony	
justice		verdict	conspiracy
#legal	#appeal	violation	insurrection
rights	#case	witness	rebellion
unconstitutional	#convict		revolt
	crime	court	revolution
alien	decision	jurisdiction	secession

*This symbol is used to indicate those words that the author would now, on the basis of further research, omit from a list of essentials.

#This symbol is used to indicate those words that were, for one reason or another, not included in the list, although their essential character is clear.

TESTS AND MEASUREMENTS

APPENDIX A: Words Selected as the Essential Core of Vocabulary in History—Continued

citizen	execution	supreme court	sedition
exile	injunction		smuggling
	judge	bribery	treason
		corruption	
F. Military Terms:			
allies	navy	draft	invasion
*belligerents	#officer	mobilization	*massacre
enemy	#reinforcements		military
hostile	#service	*aggression	*munitions
*pirate	#troops	#attack	*occupation
	submarine	#battle	offensive
#army		blockade	siege
#commander	marine	*bombardment	*strategic
confederate	militia	campaign	surrender
cruiser	recruit	contraband	#victory
#general	#soldier	defensive	#war
fleet	veteran	embargo	
#forces	volunteer	*evacuation	armistice
naval		*fortification	disarmament
			indemnity
			reparations
G. Geographical Terms:			
boundary	#district	#exploration	agriculture
continent	#region	navigation	irrigation
*continental	#section	voyage	reclamation
coast			
frontier	discover	conservation	*Pan-American
prairie	expedition	natural resources	
H. Religious Terms:			
clergy	*denomination	catholicism	papacy
missionary	*heresy	protestantism	pope
	#intolerance		
creed	persecution	crusade	
	tolerance		
I. Terms Referring to Chronology and Records:			
ancient		#history	civilized
century	era	#records	primitive
current	event		
decade	modern	propaganda	#movement
medieval	period	#publicity	precedent
			tradition

*This symbol is used to indicate those words that the author would now, on the basis of further research, omit from a list of essentials.

#This symbol is used to indicate those words that were, for one reason or another, not included in the test, although their essential character is clear.

APPENDIX B

Form A

TEST OF CONCEPTS USED IN THE SOCIAL STUDIES

Devised by L. C. PERRY, The Ohio State University

Name _____ Age _____ Grade _____

School _____ Place _____ Date _____

- * 1. Who is the head of the United States government? (a) the Governor (b) the Secretary (c) the President (d) the King
- * 2. What name is given to land next to the sea? (a) the mountains (b) the coast (c) the plains (d) the valleys
- * 3. What is the daughter of a king called? (a) a queen (b) a prince (c) a gentleman (d) a princess
- * 4. What name is given to a place where thousands of people live? (a) a house (b) a city (c) a village (d) a school
- * 5. Where would you expect to find a judge when he is at work? (a) in a courtroom (b) in a store (c) in a schoolroom (d) on a farm
6. What name is given to additions to the constitution? (a) doctrines (b) amendments (c) assemblies (d) declarations
7. Which word means loyalty to one's country? (a) neutrality (b) migration (c) patriotism (d) constitution
8. What is the law-making body of every state in the United States called? (a) parliament (b) congress (c) judiciary (d) legislature
9. In which form of government do the people have the greatest power? (a) an absolute monarchy (b) a republic (c) an oligarchy (d) an embassy
10. What term is used in referring to the entire system of courts? (a) legislative (b) administrative (c) supervision (d) judicial
11. What are people when they first enter a foreign country? (a) citizens (b) laborers (c) aliens (d) peasants
12. Which group contains the official advisers to the president? (a) House of Representatives (b) Senate (c) Supreme Court (d) Cabinet
13. What is one important purpose of politics? (a) to eliminate graft (b) to elect the best possible men (c) to organize public opinion (d) to keep candidates from being selfish
14. What is a primary? (a) a method of choosing the men to be voted on (b) a political convention (c) a group of foreign officials (d) a law dealing with capital crimes
15. What is a "ticket"? (a) the candidate who has been nominated (b) a president before election (c) a list of people to be voted on (d) an unwritten law
16. Which word refers to the affairs relating to one's own country? (a) foreign (b) international (c) domestic (d) diplomatic
17. What is meant by the "authorities"? (a) high officials (b) foreigners (c) heads of corporations (d) inventors

* USED AS EXAMPLES

TESTS AND MEASUREMENTS

18. What does suffrage mean? (a) to suffer punishment (b) to endure suffocation (c) the right to vote (d) a method of appointment
19. Which are forms of natural resources? (a) factories (b) forests (c) ranches (d) schools
20. To which does international relations refer? (a) relations within a country (b) relations between a country and its possessions (c) relations between two dependent states (d) relations between an independent state and a foreign country
21. What is an individual called when he has announced his intention of running for a certain office? (a) an opponent (b) a delegate (c) a candidate (d) an officer
22. From what are precedents often derived? (a) tradition (b) cabinet (c) temperature (d) campaigns
23. Which word means to give another person the power to do something? (a) enforce (b) refuse (c) authorize (d) excommunicate
24. Which word means opposed to slavery? (a) pro-slavery (b) ante-bellum (c) pan-American (d) anti-slavery
25. When does a political campaign take place? (a) just before a session of congress (b) just before a war (c) just before an election (d) just before a law is passed
26. What is a federal government? (a) a union of states (b) a constitution (c) a realm (d) a province
27. Which religion is chiefly opposed to Catholicism? (a) Hinduism (b) Atheism (c) Protestantism (d) Communism
28. Which are parts of the constitution? (a) articles (b) treaties (c) vetoes (d) verdicts
29. What does one nation sometimes make to another? (a) a commodity (b) a census (c) a constitution (d) a proposal
30. Of what is a governor the head? (a) a country (b) a city (c) a state (d) a town
31. What is another name for a state? (a) a commonwealth (b) a congress (c) a department (d) a city
32. What is a plank? (a) a ballot (b) a political slogan (c) an item in the platform (d) a deadlock in the party
33. What word means much the same as freedom? (a) sovereignty (b) tyranny (c) pacifism (d) liberty
34. Which are enacted? (a) verdicts (b) debts (c) laws (d) dispatches
35. Which is a recruit? (a) a profiteer (b) a banker who loans money to the enemy (c) a newly-enlisted soldier (d) a criminal who is awaiting execution
36. What is the chief business of the executive department of the government? (a) to abolish laws (b) to make laws (c) to interpret laws (d) to enforce laws
37. Which word is most similar in meaning to despotism? (a) anarchy (b) tyranny (c) confederation (d) union
38. By what process is cloth made from wool? (a) mining (b) farming (c) dyeing (d) manufacturing
39. What is another name for "goods"? (a) merchandise (b) business (c) factories (d) mines
40. Which phrase refers to the entire business of making woolen clothing? (a) the woolen shops (b) the woolen industry (c) the wool-growers (d) the sheep ranchers

ACTUAL TESTS BLANKS

41. What happens when money depreciates? (a) it becomes less valuable (b) it will buy more (c) it has to go back to the mint (d) it can be used in foreign countries
42. Who is the official head of a monarchy? (a) king (b) president (c) dictator (d) judge
43. Which is a raw material? (a) cloth (b) wood (c) paper (d) chair
44. What is the purpose of diplomacy? (a) to carry on affairs within a country (b) to elect governmental officers (c) to conduct relations between countries (d) to increase industrial output
45. What name is given to things brought into the United States from a foreign country? (a) customs (b) imports (c) pioneers (d) bonds
46. What does religious tolerance mean? (a) making all people belong to one church (b) believing everything in the Bible (c) believing in science instead of the church (d) allowing people to believe what they will
47. What is the name given to the activities of any group for the purpose of getting a certain bill through the legislature? (a) lobbying (b) electioneering (c) voting (d) secret service
48. What were the people called who wished to do away with slavery? (a) prohibitionists (b) suffragists (c) abolitionists (d) monarchists
49. What is meant by legislation? (a) the passing of laws (b) the convicting of criminals (c) the collecting of money (d) the enforcing of laws
50. What happens when a law is nullified? (a) it is enacted (b) it is ratified (c) it is no longer in force (d) it becomes a decree
51. Which word means those who enter a country to live there? (a) middleclass (b) immigrants (c) patriots (d) business men
52. Which word means giving money to an officer of the government to influence his actions? (a) bribery (b) counterfeiting (c) piracy (d) politics
53. Which laws refer to the manufacture, sale, or transportation of liquor? (a) sedition laws (b) naturalisation laws (c) prohibition laws (d) corporation laws
54. Which word means a prolonged attempt to conquer some fortified place? (a) a battle (b) a siege (c) a skirmish (d) a retreat
55. Which means a robber on the seas? (a) bandit (b) thief (c) pirate (d) highwayman
56. What is one way of attacking a fortification? (a) by skirmishing (b) by massacre (c) by reinforcements (d) by bombardment
57. What is a court's decision called? (a) a tariff (b) a verdict (c) a proposal (d) a prohibition
58. Which word means to gain someone's good will after it has once been lost? (a) tolerate (b) accuse (c) conciliate (d) condemn
59. Which is the law-making body of a city? (a) council (b) association (c) congress (d) senate
60. What word means bringing goods into a country without paying duty? (a) scandal (b) graft (c) counterfeiting (d) smuggling
61. Which word means to give each person whatever he deserves? (a) charity (b) divorce (c) despotism (d) justice
62. Which word means a stopping of war long enough to discuss peace? (a) disarmament (b) insurrection (c) treaty (d) armistice
63. Which is a result of independence, in a governmental sense? (a) unlimited freedom (b) self-government (c) control by others (d) lack of initiative

TESTS AND MEASUREMENTS

64. What is a person owned by another called? (a) an emperor (b) a savage (c) a priest (d) a slave
65. Which is an invention? (a) peninsula (b) steamboat (c) lightning (d) settlement
66. To what is a panic most often due? (a) too much capital (b) too little bankruptcy (c) too much speculation (d) too little trade
67. What is the process of carrying things from one place to another called? (a) translation (b) transition (c) transaction (d) transportation
68. What do you cast at an election? (a) a policy (b) an official (c) a promise (d) a ballot
69. What name is frequently given to a new town in a new country? (a) settlement (b) metropolis (c) prairie (d) siege
70. What does a person do when he exploits others? (a) he is a diplomat (b) he charges too much for what he sells (c) he leads a riot (d) he hires others to work for him
71. What word means an agreement between two countries to give each other trade privileges? (a) reciprocity (b) emancipation (c) finance (d) conspiracy
72. Which term means legal punishment by death? (a) assassination (b) execution (c) spy (d) imprisonment
73. When is there usually a period of reconstruction? (a) after a period of prosperity (b) after a war (c) after a religious persecution (d) after a new president is elected
74. What does "jurisdiction" mean? (a) to obey (b) to refuse loyalty to (c) to defend (d) to have power over
75. Which name is given to the head of a ministry? (a) premier (b) governor (c) consul (d) ambassador
76. Which of the following may be described as an event? (a) interest in education (b) movement toward stricter law enforcement (c) development of an industry (d) declaration of war
77. What is the money which the government collects for public use called? (a) revenue (b) appropriation (c) currency (c) allotment
78. Which belongs to the navy? (a) tank (b) regiment (c) cavalry (d) cruiser
79. Which is the final act in electing a president? (a) voting (b) petition (c) compromise (d) nomination
80. Which is a political party? (a) Puritan (b) Republican (c) Episcopal (d) Gentile
81. Which action closes ports to navigation? (a) naval battle (b) embargo (c) civil war (d) aviation
82. Which is a statement of religious beliefs? (a) creed (b) petition (c) resolution (d) treaty
83. Which is a citizen-army called out only in time of emergency? (a) volunteers (b) militia (c) brigade (d) police
84. Which political party is opposed to private ownership? (a) Anti-Saloon League (b) revolutionist (c) loyalist (d) socialist
85. What are the representatives at political conventions called? (a) commissioners (b) delegates (c) visitors (d) democrats

TEST OF CONCEPTS USED IN THE SOCIAL STUDIES

Devised by L. C. FURNER, The Ohio State University

Name _____ Age _____ Grade _____

School _____ Place _____ Date _____

1. When two countries make an agreement to help each other what is it called?
(a) competition (b) disarmament (c) alliance (d) election
2. If fifty people vote unanimously for a certain officer, how many votes did he get?
(a) 25 (b) 49 (c) 0 (d) 50
3. What is one way of getting soldiers? (a) by confiscation (b) by arrest (c) by draft (d) by pension
4. Which term describes property that represents a large money value? (a) poverty (b) wealth (c) profit (d) supply
5. With what is finance chiefly concerned? (a) farms (b) money (c) people (d) rights
6. Which is the science that is concerned particularly with ships (a) militarism (b) chemistry (c) agriculture (d) navigation
7. What is meant by reclamation? (a) civilizing a savage (b) educating a child (c) converting a heathen (d) making a waste land usable
8. Which word means an official written statement? (a) empire (b) document (c) conference (d) committee
9. Which is the opposite of a military occupation? (a) ammunition (b) evacuation (c) proscription (d) rehabilitation
10. Which word refers to those people who are employed in mills, factories, or mines?
(a) monopoly (b) wealth (c) capitalism (d) labor
11. What is given to a people to prove their ownership of the land they settle upon?
(a) petition (b) charter (c) doctrine (d) amendment
12. Which people are hostile? (a) citizens (b) employees (c) enemies (d) allies
13. What does secession mean? (a) to combine (b) to cooperate (c) to withdraw (d) to capture
14. How long is a century? (a) 10 years (b) 50 years (c) 100 years (d) 1000 years
15. Which word means the period during which a legislature is meeting? (a) defeat (b) epoch (c) session (d) congress
16. Which is a form of currency? (a) a pound of sugar (b) a college degree (c) a museum (d) a silver dollar
17. What are people permanently located in a country called? (a) inhabitants (b) foreigners (c) emigrants (d) capitalists

TESTS AND MEASUREMENTS

18. Which of the following is a process by which the people may express themselves directly on some question? (a) bureaucracy (b) paternalism (c) referendum (d) propaganda
19. What action is taken on a treaty if those voting approve it? (a) it is consumed (b) it is dissolved (c) it is ratified (d) it is annulled
20. What name is given to the chief legislative body of England? (a) parliament (b) congress (c) estates-general (d) reichstag
21. Who succeeds a president dying in office? (a) the secretary of state (b) the speaker of the House (c) the chief justice of the supreme court (d) the vice-president
22. Which word is most similar in meaning to trust? (a) mortgage (b) nation (c) justice (d) corporation
23. Who are belligerents? (a) the neutral countries (b) the countries that are defeated (c) the countries that win (d) the countries that are fighting
24. Which word means freedom from slavery? (a) pardon (b) agitation (c) emancipation (d) exemption
25. Which word means privileges no one can legally take from us? (a) laws (b) doctrines (c) rights (d) institutions
26. Who may be impeached? (a) a business man (b) a criminal (c) a reporter (d) a president
27. Which are called clergy? (a) ministers (b) statesmen (c) governors (d) judges
28. Which word means freedom from disturbance? (a) treaty (b) peace (c) hostility (d) law
29. On what is there a tariff? (a) things manufactured in a country (b) things sent out of a country (c) things invented in a country (d) things sent into a country
30. Which is a way of filling certain public offices? (a) application (b) resignation (c) endorsement (d) appointment
31. What is a poll? (a) a place to vote (b) a ballot (c) an unfair election (d) a type of violence
32. Which word is similar in meaning to invalid? (a) delinquent (b) unconstitutional (c) fugitive (d) immoral
33. What determines the necessities, comforts, and luxuries which the average home insists upon? (a) public opinion (b) states' rights (c) standard of living (d) universal suffrage
34. Which word refers to political corruption? (a) graft (b) lynching (c) revolt (d) mutiny
35. What is the money called that one country borrows from another (a) debt (b) a dividend (c) a premium (d) a treaty
36. When is a person an exile? (a) when he is sick (b) when he holds an office (c) when he has to live in a country not his own (d) when he has been in prison
37. What word might we use in place of illegal? (a) injustice (b) liberty (c) illegality (d) unlawful
38. What is at the head of the judiciary branch of the government (a) public opinion (b) the district attorney (c) capital punishment (d) the supreme court

ACTUAL TESTS BLANKS

39. What are those people called who are living in the country they were born in? (a) civilians (b) citizens (c) Americans (d) aliens
40. What is a confederacy? (a) a decree (b) a league (c) a treaty (d) a petition
41. What do you call a more or less permanent settlement in a new country? (a) a democracy (b) a colony (c) a tariff (d) a continent
42. What is repudiated? (a) debts (b) governors (c) mortgages (d) land
43. What is a caucus for? (a) to elect a candidate (b) to defeat a candidate (c) to decide upon a candidate (d) to punish an unwilling candidate
44. Of what bodies is Congress composed? (a) Parliament and Assembly (b) Prime Minister and Cabinet (c) House of Representatives and Senate (d) House of Lords and House of Commons
45. Which is the name of the army, in American history, that fought for the southern states? (a) English (b) confederate (c) union (d) colonial
46. What is the title of a member of the higher branch of the Congress of the United States? (a) representative (b) senator (c) delegate (d) judge
47. Which is another word for commerce? (a) government (b) trade (c) religion (d) crime
48. What are the views of people in general called? (a) trade unions (b) public opinion (c) states' rights (d) civil service
49. What does a company sell when it wants to increase its capital? (a) salaries (b) drafts (c) fees (d) stocks
50. What does interstate mean? (a) within a state (b) between cities (c) between states (d) within a country
51. What is the official ceremony that marks the beginning of the President's term called? (a) nomination (b) coronation (c) convention (d) inauguration
52. What do we call the place where disagreements are tried before a judge? (a) capital (b) navy (c) court (d) public opinion
53. What happens when money is inflated? (a) it will buy more (b) it is not possible to invest it (c) it can be used to pay international debts (d) it will buy less
54. Which is the flat or rolling type of country? (a) seashore (b) valley (c) prairie (d) divide
55. Of what is a "grant" often composed? (a) land (b) people (c) labor (d) officers
56. What name is given to an area of partly settled land belonging to a country? (a) province (b) nation (c) republic (d) investment
57. What does majority mean? (a) a minority (b) an officer in the army (c) more than fifty per cent of the votes (d) a successful candidate
58. What name is given to the President's term of office? (a) decade (b) panic (c) campaign (d) administration
59. What name is given to the money or goods paid by one country to another as settlement for damage done during a war? (a) tariff (b) currency (c) reparations (d) loan
60. To which does naval refer? (a) ships (b) deserts (c) trains (d) tunnels

TESTS AND MEASUREMENTS

61. By what method is the governor's office filled? (a) by initiative (b) by election (c) by recall (d) by endorsement
62. Who occupies the papacy? (a) the pope (b) the president (c) the prime minister (d) the police
63. What are the people called who join an army of their own accord? (a) generals (b) aliens (c) capitalists (d) volunteers
64. Who decides whether or not a man is guilty of murder? (a) the jury (b) the judge (c) the witnesses (d) the lawyers
65. What is made in the mint? (a) candy (b) money (c) clothes (d) bacon
66. What is the outer edge of a civilized area called? (a) metropolis (b) suburbs (c) frontier (d) seacoast
67. Of whom is the population composed? (a) banks (b) people (c) debts (d) proposals
68. When is an injunction most often used? (a) during a strike (b) during a battle (c) during an exploration (d) during a voyage
69. What is a tariff which seeks to help home industries? (a) tariff for revenue only (b) international tariff (c) free trade (d) protective tariff
70. What is the opposite of "urban"? (a) foreign (b) metropolitan (c) illiterate (d) rural
71. Which word means open revolt against authority? (a) diplomacy (b) agriculture (c) debate (d) insurrection
72. Which refers to the lowest social class in many countries? (a) representatives (b) peasants (c) inhabitants (d) delegates
73. What are nations on the same side in a war called? (a) officers (b) allies (c) marshals (d) civilians
74. Which is a mob? (a) a fleet of battleships (b) a decrease in prices (c) an excited group of people (d) a tax on imports
75. Which is a form of conspiracy against one's country? (a) sedition (b) opposition (c) criticism (d) immigration
76. When a law is repealed, what is always the result? (a) it is enforced more rigidly (b) it is amended (c) it is no longer in force (d) it is a source of dissatisfaction
77. What is the line that shows where one country stops and another begins? (a) boundary (b) monument (c) compass (d) ocean
78. Which is something to be obeyed? (a) law (b) proposition (c) report (d) celebration
79. Which term means the closing of the ports of one country by the naval forces of another? (a) embargo (b) naval battle (c) decisive (d) blockade
80. How does a person become a candidate for office? (a) by defeat (b) by nomination (c) by election to office (d) by appointment

TEST OF CONCEPTS USED IN THE SOCIAL STUDIES*Devised by L. C. PRESSY, The Ohio State University*

Name..... Age..... Grade.....

School..... Place..... Date.....

- * 1. Who of the following represents the United States officially in a foreign country?
(a) an ambassador (b) a senator (c) a patriot (d) a governor
- * 2. What is a group of people carrying on a particular business called? (a) a mandate (b) a company (c) dividends (d) a community
- * 3. What is likely to result when all parties in the government have equal strength?
(a) imperialism (b) massacre (c) amendment (d) deadlock
- * 4. Where did the rich people of the South live in the early days? (a) in homesteads (b) in log cabins (c) on plantations (d) on river boats
- * 5. What word means about the same as revolution? (a) extradition (b) administration (c) penalty (d) rebellion
- 6. What is necessary if desert land is to become fruitful? (a) migration (b) reparations (c) irrigation (d) capital
- 7. To what does agriculture refer? (a) fishing (b) mining (c) farming (d) manufacturing
- 8. What is the opposite of "modern"? (a) mediaeval (b) colonial (c) ancient (d) foreign
- 9. Which is an expression of the government's opinion on something important?
(a) doctrine (b) empire (c) budget (d) commerce
- 10. Which term refers to a religious war? (a) an expansion (b) a revolution (c) a crusade (d) an offensive
- 11. What term may be used to describe the peaceful settlement of disagreements?
(a) arbitration (b) annexation (c) nullification (d) proclamation
- 12. What is the nation's capital? (a) the largest city (b) the chief seaport (c) the seat of its government (d) the center of its wealth
- 13. How long is a decade? (a) 10 years (b) 25 years (c) 50 years (d) 100 years
- 14. When do current events take place? (a) a long time ago (b) a short time ago (c) in the future (d) in the present
- 15. Which name is given to the countries ruled by an emperor? (a) state (b) alliance (c) commonwealth (d) empire
- 16. What kind of people abdicate? (a) traitors (b) employees (c) kings (d) slaves
- 17. Which is a denomination? (a) Baptist (b) Jew (c) German (d) Negro

* USED AS EXAMPLES

TESTS AND MEASUREMENTS

18. What do two nations have to do if they want quite different things but do not want to go to war about their differences? (a) repeal (b) ratify (c) compromise (d) mobilize
19. What are the dealings of one country with another called? (a) amendments (b) documents (c) restorations (d) negotiations
20. Which adjourns? (a) a treaty (b) the king (c) congress (d) the constitution
21. Which is an old soldier? (a) an invalid (b) a minister (c) a veteran (d) a convict
22. What is usually the object of an expedition? (a) to invent something (b) to control something (c) to plan something (d) to discover something
23. Toward which is the conservation movement directed? (a) toward political corruption (b) toward business reform (c) toward religious disputes (d) toward preserving forests
24. Which is a continent? (a) Panama (b) London (c) Asia (d) France
25. What usually determines your nationality? (a) the people you live with (b) the school you attend (c) the country you were born in (d) the government of your city
26. How does the navy get money to build a new ship? (a) by annulment (b) by appropriation (c) by acquisition (d) by invention
27. Which is a method of deciding how much tax should be paid by different people? (a) assessment (b) endorsement (c) adjournment (d) annulment
28. What is another name for the United States government? (a) the Confederacy (b) the Union (c) the President (d) the Conference
29. In the early days of the United States which party favored a strong central government? (a) Tories (b) Know-nothings (c) Federals (d) Democrats
30. Where is the government's money kept? (a) in the House (b) in the Senate (c) in the mint (d) in the treasury
31. What does a city ordinarily sell when it needs money for some public improvement? (a) credits (b) lands (c) dividends (d) bonds
32. Where are customs officers found? (a) in cities of entry (b) in state capitals (c) in sweat shops (d) in the ten largest cities
33. Which word means a method of influencing people to agree with you? (a) revolt (b) propaganda (c) election (d) victory
34. What is plotting against one's own country called? (a) murder (b) witchcraft (c) riot (d) treason
35. Which is a small political division for local government? (a) section (b) region (c) county (d) territory
36. What do you do when you leave your own country to settle in another? (a) emigrate (b) retreat (c) surrender (d) abdicate
37. Which is annexed? (a) revolt (b) credits (c) land (d) neutrality
38. What does the President sometimes appoint when he wants some matter investigated? (a) a company (b) an embassy (c) a commission (d) a delegate
39. What gives the people the "fundamental law" by which they are governed? (a) constitution (b) manifesto (c) assembly (d) parliament

ACTUAL TESTS BLANKS

40. Which is a law? (a) a petition (b) a dispatch (c) a census (d) a decree
41. What is a king sometimes called? (a) chancellor (b) prime minister (c) president (d) sovereign
42. Which party was opposed to England during the Revolutionary War? (a) Whig (b) Progressive (c) Abolitionist (d) Tory
43. What is a person called who locates in an unsettled section of his own country? (a) emigrant (b) pioneer (c) slave (d) proprietor
44. What name is given to the organization of employees? (a) nations (b) political parties (c) pioneers (d) unions
45. What word means a place where people live? (a) a monopoly (b) a community (c) a court (d) a decade
46. What are affairs within a country called? (a) restricted (b) financial (c) manufacture (d) internal
47. What is a civil war? (a) a war during the past 25 years (b) a war involving several nations (c) a war between two nations (d) a war within a single country
48. What are the strongest nations in the world often called? (a) empires (b) leagues of nations (c) powers (d) democrats
49. What is the opposite of offensive? (a) conservative (b) active (c) inventive (d) defensive
50. Which word is most similar in meaning to period? (a) year (b) season (c) month (d) era
51. What happens when there is a financial crisis? (a) a new president is elected (b) prices are lowered (c) there is a war (d) there is a period of prosperity
52. What is a veto? (a) a compromise (b) a political party (c) a demand for funds (d) a refusal by the president
53. What is the tax put upon things brought into a country? (a) bonus (b) duty (c) discount (d) income
54. Which word refers to things done by soldiers? (a) military (b) naval (c) civil (d) aeronautical
55. What is the act of calling out troops called? (a) mobilization (b) confirmation (c) indictment (d) recantation
56. Which is a written law? (a) a negotiation (b) a statute (c) a publication (d) a veto
57. If you and someone else both want to be elected President, what do you call the other person? (a) opponent (b) ally (c) vice-president (d) democrat
58. What is the largest group of people living under the same government? (a) city (b) nation (c) county (d) province
59. What is the result of a restriction? (a) you refuse to do something (b) you are prevented from doing something (c) you are encouraged to break a law (d) you are put in jail
60. Which name is given to the misuse of patronage? (a) civil service (b) spoils system (c) factory system (d) secret service
61. What policy is opposed to that of centralization in government? (a) states' rights (b) competition (c) public opinion (d) magna charta

TESTS AND MEASUREMENTS

62. What do we call the point of view of a person strongly in favor of one side of an argument? (a) insane (b) dominating (c) victorious (d) partisan
63. Which word implies the opposite of primitive? (a) early (b) civilized (c) ancient (d) original
64. What is given by a railroad to certain companies so they can send things more cheaply than other companies? (a) resolution (b) rebate (c) ballot (d) competition
65. What does a broker buy in the "market"? (a) banks (b) stocks (c) merchandise (d) labor
66. Which of the following words applies to the practice of lowering of prices by one company so as to get business away from another? (a) rebellion (b) emigration (c) output (d) competition
67. Which word means to send goods out of the country in which they are produced? (a) retail (b) export (c) sell at wholesale (d) import
68. What is the exclusive right or privilege of engaging in a particular business called? (a) competition (b) monopoly (c) capital (d) research
69. Which is a method by which a government raises money? (a) rebates (b) diplomacy (c) taxes (d) salaries
70. Who of the following go to foreign lands to teach religion (a) converts (b) missionaries (c) ambassadors (d) bishops
71. Which is a commodity? (a) wheat (b) people (c) scenery (d) education
72. What is usually invested? (a) a mortgage (b) some money (c) a person (d) some goods
73. What name is given to a person who gives testimony before a court? (a) witness (b) lawyer (c) judge (d) juror
74. To what places does Pan-American refer? (a) U. S. and Canada (b) America and Europe (c) North and South America (d) the U. S. and its possessions
75. Which word means any commercial or industrial undertaking? (a) banking (b) business (c) exchange (d) contract
76. What is the condition when there is no government? (a) slavery (b) freedom (c) expansion (d) anarchy
77. Which word means that a country has taken over control of a smaller country without the latter's consent? (a) imperialism (b) democracy (c) anarchy (d) patriotism
78. Which is a diplomatic agent? (a) a president (b) a jury (c) a minister (d) a priest
79. What is a group of ships under one command called? (a) fleet (b) caravan (c) frigate (d) army
80. Which word describes any country in which a person is not a citizen? (a) foreign (b) native (c) domestic (d) continental

TEST OF CONCEPTS USED IN THE SOCIAL STUDIES

Devised by L. C. FRENCH, The Ohio State University

Name _____ Age _____ Grade _____

School _____ Place _____ Date _____

1. What name was given to large areas before they became states? (a) cities (b) factories (c) territories (d) jungles
2. What is a budget? (a) a cash account (b) a list of names (c) an estimate of expenditures (d) a payroll
3. What happens when a law is put into effect? (a) it is legislated (b) it is decreed (c) it is sanctioned (d) it is enforced
4. What is the name of a formal agreement between two or more countries? (a) a treaty (b) an edict (c) a resolution (d) a statute
5. Which is a journey on the ocean? (a) caravan (b) conspiracy (c) voyage (d) treason
6. Which boat can travel under the water? (a) cruisers (b) dreadnaught (c) iron-clad (d) submarine
7. Who are the "soldiers of the sea"? (a) the sailors (b) the militia (c) the marines (d) the population
8. Which word means some dishonest act? (a) abdication (b) prohibition (c) suffrage (d) fraud
9. Which is a war of aggression? (a) a lockout (b) an invasion (c) a retreat (d) a battle
10. Which adjective refers to the country in which one is born? (a) alien (b) foreign (c) adopted (d) native
11. Which is a public statement? (a) proclamation (b) imperialism (c) exposition (d) assembly
12. Which word means to give approval? (a) repeal (b) sanction (c) rule (d) revolt
13. What is meant by democracy? (a) government by a king (b) government by a dictator (c) government by the people (d) government by the nobility
14. When an army has lost a war what does it do? (a) it dies (b) it surrenders (c) it conquers (d) it maneuvers
15. What group of people rank themselves highest in society? (a) middlemen (b) aristocrats (c) pioneers (d) democrats
16. What does a census usually count? (a) resources (b) capital (c) expenditures (d) people

TESTS AND MEASUREMENTS

17. What does migration mean? (a) a primitive society (b) a kind of law (c) a journeying from place to place (d) a form of industry
18. Which term may refer to a division within a department of the government? (a) bureau (b) court (c) congress (d) cabinet
19. What is a plan for several battles called? (a) siege (b) navy (c) campaign (d) expansion
20. When a government tries to improve conditions, what is it called? (a) referendum (b) initiative (c) reform (d) campaign
21. What must you have in order to borrow money? (a) bills (b) lawyers (c) credit (d) corporation
22. Which of the following is a method of communication? (a) investment (b) telephone (c) production (d) politics
23. Which of the following statements best describes a bankrupt country? (a) it has too many people in it (b) it has no agriculture (c) it cannot pay its debts (d) it has lost its colonies
24. Which of the following is a legislative assembly? (a) empire (b) congress (c) world court (d) bureau of education
25. How do laborers most often attempt to force an employer to raise their wages? (a) by going to war (b) by going to another factory (c) by going to a foreign country (d) by going on strike
26. What does it mean to be naturalized? (a) to go to another country (b) to hold an office in a government (c) to become a citizen of a foreign country (d) to become a very wealthy person
27. What name is given to a small organized group of persons with certain work to do? (a) a committee (b) an assemblage (c) a parliament (d) a congress
28. Which are "public utilities"? (a) gas companies (b) state banks (c) cotton mills (d) iron foundries
29. What name is given to the money used for financing industry? (a) tax (b) treasury (c) regime (d) capital
30. What is money paid out by the government called? (a) expenditures (b) taxes (c) tariffs (d) assessments
31. To what expression does the term coinage refer? (a) the weaving of cloth (b) the raising of wheat (c) the mining of metals (d) the making of pieces of money
32. What is the money called which the loser of a war pays as damages to the winner (a) loan (b) appropriation (c) mortgage (d) indemnity
33. Which are munitions? (a) soldiers (b) uniforms (c) transports (d) shells
34. Which term refers to cities? (a) national (b) interstate (c) municipal (d) provincial
35. What is a platform? (a) a list of candidates (b) a statement of principles (c) a religious creed (d) a political slogan
36. What is a petition? (a) a reply (b) a refusal (c) a repeal (d) a request
37. In what period were battles carried on by knights in armor? (a) the colonial period (b) the pre-historic period (c) the mediæval period (d) the revolutionary period

ACTUAL TESTS BLANKS

36. What term explains the position of a country which does not take sides in a war between two other countries? (a) democratic (b) honest (c) partial (d) neutral
39. Which are members of the House? (a) delegates (b) senators (c) mayors (d) representatives
40. Which word is used to describe some important problem that needs to be settled? (a) an issue (b) a decision (c) an edict (d) a sanction
41. What is the opposite of conservative? (a) social (b) radical (c) political (d) illegal
42. Which of the following adjectives best describes a law that is based not on any good reasons but on personal likes or dislikes? (a) democratic (b) illegal (c) foreign (d) arbitrary
43. When a legislature has voted for something what is it called? (a) an idea (b) an act (c) a crime (d) a proposal
44. What does violation mean? (a) breaking a law (b) signing a treaty (c) electing a candidate (d) declaring a war
45. Which usually has the smallest number of people in it? (a) nation (b) state (c) city (d) town
46. Which word means the killing of a large number of people? (a) an encounter (b) an economy (c) a decree (d) a massacre
47. Which word describes a condition in which people have more money than usual? (a) poverty (b) prosperity (c) prohibition (d) protection
48. To what does the word "greenback" refer? (a) a division of an army (b) a kind of ship (c) a kind of money (d) a department of the treasury
49. What is the user of goods called? (a) the consumer (b) the reformer (c) the traitor (d) the producer
50. Which are contraband during a war? (a) houses (b) guns (c) battles (d) islands
51. What do nations do when they disagree, but want to find some friendly way out of their difficulties? (a) declare war (b) hold a conference (c) get out propaganda (d) elect new officials
52. Which is a representative of the government in a foreign country? (a) consul (b) sovereign (c) vice-president (d) banker
53. How did the early Americans state their independence from England? (a) by a riot (b) by a treaty (c) by a political party (d) by a declaration
54. In American history what name was given to a part of the country before it became a state? (a) republic (b) territory (c) statute (d) nation
55. What is Canada called? (a) a frontier (b) a dominion (c) a despotism (d) an empire
56. What name is given to the giving up of armies and navies? (a) compromise (b) caucus (c) speculation (d) disarmament
57. When a fort is especially well placed what is it called? (a) strategic (b) colonial (c) conservative (d) patriotic
58. By whom are examinations held for the filling of many positions in the government? (a) party politics (b) public opinion (c) League of Nations (d) civil service

TESTS AND MEASUREMENT.

59. Which word means to give in to some other persons's ideas? (a) enforce (b) nullify (c) concede (d) veto
60. What name is given to the movement of people from the eastern parts of the United States to the western? (a) secession (b) expansion (c) conservation (d) precedent
61. What are the heads of departments in the United States government called? (a) consuls (b) premiers (c) secretaries (d) senators
62. What do you do when you pass a resolution? (a) vote (b) nullify (c) annex (d) veto
63. Which is the name of the houses lived in by some of the early settlers in the Middle West? (a) hotels (b) homesteads (c) plantations (d) bungalows
64. When a group of people get very angry and excited and want to show how they feel what sometimes takes place? (a) an appropriation (b) a riot (c) a campaign (d) monopoly
65. Which word refers to some certain period of time in history? (a) politics (b) neutrality (c) regime (d) massacre
66. Which word indicates the government's attitude on some important question? (a) policy (b) issue (c) patronage (d) crisis
67. What exists when one group of people are unfair and cruel to another group? (a) independence (b) tolerance (c) treason (d) oppression
68. For what have people often been persecuted by the church? (a) migration (b) war (c) heresy (d) contraband
69. Which party is at present opposed to the Republicans? (a) Democratic (b) Federalist (c) Liberal (d) Abolitionist
70. Which term refers to the process of making things in an industry? (a) production (b) raw material (c) wages (d) resources
71. Which word means something to be voted on? (a) bill (b) credit (c) diplomacy (d) proclamation
72. Which party developed within the Republican party during the past twenty years? (a) Whig (b) Democratic (c) Progressive (d) Socialist
73. Which word refers to the conditions in the place where you live? (a) religious (b) universal (c) national (d) local
74. Which term is applied to a person who holds a position of authority in the government? (a) principal (b) lawyer (c) official (d) servant
75. Which is a race? (a) French (b) Negro (c) American (d) Quaker
76. When a government wishes to punish people, what does it often do with their property? (a) defends it (b) confiscates it (c) subordinates it (d) makes it into forts
77. Which word means the breaking of the law? (a) strikes (b) war (c) crime (d) wealth
78. What is an assembly? (a) the allies in a war (b) a group of people (c) a high tariff (d) a set of laws
79. Which word means the constituted authorities of a nation? (a) the banks (b) the president (c) the government (d) the supreme court
80. Which is a "continental" country? (a) Africa (b) Mexico (c) Canada (d) France

APPENDIX C

APPENDIX C: Distributions of Total Scores on Preliminary Forms B, C, and D of the Tests

PRELIMINARY FORM B					
Score	Grades				
	4	6	8	10	12
80.....
78-79.....	1	8	40
76-77.....	6	13	91
74-75.....	5	29	103
72-73.....	16	53	141
70-71.....	35	62	134
68-69.....	45	66	112
66-67.....	..	4	65	87	105
64-65.....	..	2	68	120	69
62-63.....	..	1	76	78	38
60-61.....	..	4	111	53	41
58-59.....	..	4	123	55	25
56-57.....	1	8	126	58	20
54-55.....	..	15	91	53	14
52-53.....	3	15	99	51	8
50-51.....	..	19	86	40	7
48-49.....	..	23	85	24	5
46-47.....	1	27	74	12	8
44-45.....	2	27	52	12	2
42-43.....	..	36	49	11	2
40-41.....	3	44	49	7	2
38-39.....	3	34	25	4	2
36-37.....	7	44	34	6	1
34-35.....	3	47	25	6	2
32-33.....	11	31	10	5	2
30-31.....	17	22	12	3	1
28-29.....	24	19	7	6	..
26-27.....	31	22	10	..	1
24-25.....	20	14	2	2	..
22-23.....	22	4	1
20-21.....	13	5	5	1	..
18-19.....	7	3	3	1	..
16-17.....	11	4	1
14-15.....	10	1
12-13.....	7	2
10-11.....	3	1
8-9.....	4	1
6-7.....	3
4-5.....	1
2-3.....	2
0-1.....
Total Cases.....	209	483	1,397	926	976
Median.....	26.1	39.2	55.5	63.4	70.3

TESTS AND MEASUREMENTS

PRELIMINARY FORM C

Score	Grades				
	4	6	8	10	12
80°
78-79
76-77
74-75	4	6	121
72-73	..	1	8	32	173
70-71	8	32	140
68-69	17	46	124
66-67	23	48	117
64-65	..	1	24	61	84
62-63	24	68	73
60-61	1	2	51	37	63
58-59	..	2	61	52	32
56-57	70	62	36
54-55	..	2	70	56	37
52-53	2	8	93	49	28
50-51	..	5	94	50	21
48-49	..	16	104	44	8
46-47	1	12	108	33	14
44-45	..	12	139	23	6
42-43	2	14	111	33	5
40-41	..	30	99	18	..
38-39	1	26	107	19	4
36-37	4	25	87	10	1
34-35	..	30*	71	8	2
32-33	2	46	46	4	2
30-31	6	49	44	8	..
28-29	7	38	22	4	..
26-27	16	36	19	5	..
24-25	13	35	12	5	..
22-23	25	29	8
20-21	18	12	6	2	..
18-19	12	9	4	1	..
16-17	26	6	..	1	..
14-15	20	8
12-13	12	2
10-11	8	2
8-9	12	1
6-7	7	1
4-5	7
2-3	4
0-1	8
Total Cases.....	214	460	1,534	817	1,100
Median.....	18.3	32.0	45.9	57.2	68.2

*It should be remembered that five of the eighty items of this form were used as examples, leaving the highest possible score at 75.

APPENDIX C
PRELIMINARY FORM D

Score	Grades				
	4	6	8	10	12
80.....
78-79.....	4	11	96
76-77.....	6	14	108
74-75.....	11	19	98
72-73.....	18	30	81
70-71.....	36	26	71
68-69.....	..	1	35	46	57
66-67.....	..	1	61	52	42
64-65.....	..	3	70	33	46
62-63.....	..	2	85	56	32
60-61.....	..	2	99	45	27
58-59.....	..	3	108	43	29
56-57.....	..	6	114	42	16
54-55.....	1	13	100	39	14
52-53.....	..	10	111	27	11
50-51.....	1	19	117	25	7
48-49.....	1	15	115	15	8
46-47.....	3	28	89	22	8
44-45.....	..	31	72	10	7
42-43.....	1	36	67	9	7
40-41.....	3	29	38	6	7
38-39.....	1	20	42	1	5
36-37.....	5	29	33	2	3
34-35.....	4	30	16	2	3
32-33.....	4	33	9	2	8
30-31.....	11	33	13	2	3
28-29.....	18	31	5	3	2
26-27.....	11	19	4	..	3
24-25.....	17	18	5	1	2
22-23.....	19	7	2	..	2
20-21.....	17	16
18-19.....	24	8
16-17.....	16	6	1	..	3
14-15.....	14	3	2
12-13.....	6	1	1
10-11.....	9	1	1	..	1
8-9.....	2	1	1
6-7.....	5
4-5.....	3
2-3.....	4
0-1.....	1
Total Cases.....	201	455	1,488	583	810
Median.....	22.0	37.4	54.0	61.8	71.4

TESTS AND MEASUREMENTS

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
ambassador.....	(Used as example and not scored)					C, 1
authorities.....	21	40	81	87	94	A, 17
consul.....	19	38	61	83	87	D, 52
governor.....	61	61	84	88	97	A, 30
king.....	(Not counted because of error in scoring key)					A, 42
minister.....	11	26	32	32	51	C, 78
official.....	28	59	78	91	94	D, 74
premier.....	10	30	55	57	57	A, 75
prime-minister.....	(Not included in test)					
police.....	(Not included in test)					
president.....	(Used as example and not scored)					A, 1
representative.....	26	53	62	76	82	D, 39
secretary.....	4	21	23	30	59	D, 61
senator.....	27	49	59	60	79	D, 46
sovereign.....	32	45	54	72	86	C, 41
statesman.....	(Not included in test)					
vice-president.....	41	74	96	98	99	B, 21
assembly.....	39	61	85	85	91	D, 78
bureau.....	10	17	26	52	64	D, 18
board.....	(Not included in test)					
cabinet.....	11	17	75	84	93	A, 12
commission.....	25	34	20	31	62	C, 38
committee.....	28	80	92	98	98	D, 27
conference.....	26	75	90	92	92	D, 51
congress.....	21	62	90	94	99	B, 44
council.....	30	43	73	86	93	A, 59
department.....	4	21	23	30	59	D, 61
House of Rep.....	21	62	90	94	99	B, 44
league.....	23	32	43	71	86	B, 40
legislature.....	16	18	62	69	78	A, 8
parliament.....	7	50	83	93	93	B, 20
senate.....	21	62	90	94	99	B, 44
session.....	29	42	78	92	98	B, 15
anarchy.....	12	25	60	71	79	C, 76
*commonwealth.....	30	38	57	79	79	A, 31
communism.....	(Not included in test)					
confederacy.....	23	32	43	71	86	B, 40
despotism.....	10	11	28	36	48	A, 37
democracy.....	28	50	80	84	95	D, 13
empire.....	37	88	88	97	98	C, 15
federal.....	50	60	66	75	84	A, 26
government.....	10	25	44	51	51	D, 79

*This symbol is used to indicate those words the author would not now include in a list of common words, but upon which test data are available.

APPENDIX D

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—Continued

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
*imperialism	(Test item unsatisfactory)					C, 77
monarchy	(Not counted because of error in scoring key)					A, 42
republic	54	79	87	89	94	A, 9
self-government	37	42	76	76	85	A, 63
tyranny	10	11	28	36	48	A, 37
union	28	57	58	69	84	C, 28
city	(Used as example and not scored)					A, 4
colony	32	77	89	91	96	B, 41
country	(Not included in test)					
county	25	36	53	67	72	C, 35
*dominion	34	71	81	88	89	D, 55
nation	37	68	82	89	91	C, 58
province	14	49	78	88	94	B, 56
state	61	61	84	88	97	A, 30
territory	37	88	96	97	98	D, 1
town	81	92	97	99	99	D, 45
alliance	18	62	74	91	99	B, 1
arbitration	13	23	31	68	83	C, 11
diplomacy	13	23	47	56	78	A, 44
foreign	45	62	77	86	90	C, 80
international	13	36	25	40	60	A, 20
negotiation	25	28	42	69	87	C, 19
neutrality	21	55	84	90	94	D, 38
pact	(Not included in test)					
peace	43	72	77	81	85	B, 28
powers	19	30	36	56	76	C, 48
reciprocity	18	27	27	37	62	A, 71
treaty	34	71	91	92	97	D, 4
amendment	4	27	83	90	97	A, 6
article	18	22	69	86	96	A, 28
charter	21	58	60	65	77	B, 11
constitution	30	40	68	82	92	C, 39
act	9	28	69	69	73	D, 43
bill	11	15	59	81	84	D, 71
declaration	34	62	73	84	84	D, 53
decree	28	49	39	53	79	C, 40
document	4	47	84	96	98	B, 8
law	63	90	96	97	98	B, 78
legislation	28	41	72	79	87	A, 49
measure	(Not included in test)					
petition	21	33	55	74	89	D, 36

*This symbol is used to indicate those words the author would not now include in a list of essential terms, but upon which test data are available.

TESTS AND MEASUREMENTS

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—Continued

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
proclamation.....	17	45	60	85	93	D, 11
proposal.....	17	38	69	79	91	A, 29
provision.....	(Not included in test)					
report.....	(Not included in test)					
resolution.....	11	29	39	47	70	D, 62
restriction.....	16	28	50	73	88	C, 59
statute.....	4	15	17	52	77	C, 56
abolish.....	(Not included in test)					
*abdicate.....	12	20	27	34	54	C, 16
adjourn.....	23	48	72	89	96	C, 20
annex.....	14	29	47	65	91	C, 37
appoint.....	14	39	42	71	91	B, 30
authorize.....	11	23	60	80	92	A, 23
compromise.....	28	46	64	88	95	C, 18
concede.....	20	29	54	64	79	D, 59
*conciliate.....	30	31	56	73	91	A, 58
*confiscate.....	21	33	51	71	78	D, 76
enact.....	47	47	48	56	73	A, 34
enforce.....	39	66	88	88	88	D, 3
grant.....	(Not included in test)					
impeach.....	18	29	60	87	96	B, 26
inaugurate.....	23	24	73	90	97	B, 51
nullify.....	26	41	62	80	90	A, 50
ratify.....	12	15	54	73	83	B, 19
repeal.....	43	50	58	66	83	B, 76
*repudiate.....	12	13	28	42	55	B, 42
*sanction.....	16	28	50	73	88	D, 12
veto.....	23	41	61	88	96	C, 52
executive.....	20	27	63	65	65	A, 36
legislative.....	23	39	81	86	93	D, 24
judiciary.....	39	49	55	71	83	A, 10
appropriation.....	19	41	47	64	85	C, 26
assessment.....	23	37	66	90	95	C, 27
budget.....	13	20	76	91	96	D, 2
*coinage.....	26	62	93	97	97	D, 31
currency.....	47	51	86	94	99	B, 16
customs.....	23	46	58	73	87	C, 32
debt.....	27	74	78	83	83	B, 35
duty.....	16	27	66	80	93	C, 53
expenditures.....	15	15	43	69	82	D, 30
*greenback.....	20	37	83	93	93	D, 48

*This symbol is used to indicate those words the author would not now include in a list of essential terms, but upon which test data are available.

APPENDIX D

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—Continued

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
*mint.....	32	47	72	88	97	B, 65
protective.....	18	31	63	74	91	B, 69
revenue.....	8	19	48	54	71	A, 77
tariff.....	27	33	78	82	86	B, 29
tax.....	56	83	96	96	98	C, 69
treasury.....	73	84	95	95	95	C, 30
doctrine.....	13	29	62	87	92	C, 9
issue.....	11	13	18	30	63	D, 40
policy.....	11	24	25	39	62	D, 66
reservation.....	(Not included in test)					
centralization.....	16	29	42	43	71	C, 61
civil.....	26	45	65	79	91	C, 47
civil service.....	8	19	42	78	81	D, 58
domestic.....	11	14	31	54	75	A, 16
internal.....	23	37	49	70	90	C, 46
interstate.....	21	25	29	55	62	B, 50
local.....	26	42	71	87	87	D, 73
municipal.....	10	24	41	57	84	D, 34
states rights.....	16	29	42	43	71	C, 61
administration.....	41	45	74	79	97	B, 58
regime.....	(Item unsatisfactory, not scored)					D, 65
*capitol.....	19	73	88	90	94	C, 12
patriotism.....	20	56	92	97	98	A, 7
prohibition.....	22	45	88	94	96	A, 53
reconstruction.....	28	41	78	75	85	A, 73
referendum.....	17	20	38	52	54	B, 18
campaign.....	8	52	79	89	94	A, 25
candidate.....	15	49	82	92	97	A, 21
caucus.....	17	17	35	50	53	B, 43
convention.....	(Not included in test)					
deadlock.....	(Used as example and not scored)					C, 3
delegate.....	15	30	75	81	92	A, 85
nominate.....	25	38	71	82	88	B, 80
opponent.....	19	51	82	96	96	C, 57
plank.....	38	38	47	56	68	A, 32
platform.....	6	19	26	41	55	D, 35
politics.....	(Item unsatisfactory, not scored)					A, 13
ticket.....	26	35	59	79	84	A, 15

*This symbol is used to indicate those words the author would not now include in a list of essential terms, but upon which test data are available.

TESTS AND MEASUREMENTS

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—Continued

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
*abolitionist.....	13	18	77	83	92	A, 48
*anti-slavery.....	28	43	84	90	91	A, 24
democrat.....	33	51	87	91	92	D, 69
*federalist.....	23	40	44	48	67	C, 29
political party.....	50	59	92	88	95	A, 80
progressive.....	15	23	16	23	41	D, 72
republican.....	33	51	87	91	92	D, 69
socialist.....	11	25	42	50	75	A, 84
*whig.....	11	21	35	50	65	C, 42
majority.....	34	52	92	94	98	B, 57
minority.....	(Not included in test)					
unanimous.....	75	82	85	90	90	B, 2
ballot.....	27	43	89	92	97	A, 68
election.....	37	70	86	89	92	B, 61
polls.....	34	54	87	93	93	B, 31
primary.....	22	41	59	77	93	A, 14
suffrage.....	13	22	58	66	92	A, 18
vote.....	51	62	70	70	83	A, 79
conservative.....	9	20	31	52	78	D, 41
*partisan.....	19	37	52	57	66	C, 62
radical.....	9	20	31	52	78	D, 41
lobbying.....	2	9	24	42	63	A, 47
patronage.....	8	22	26	61	67	C, 60
spoils system.....	8	22	26	61	67	C, 60
business.....	17	43	53	65	71	C, 75
commerce.....	18	52	84	95	99	B, 47
commodity.....	7	18	18	38	79	C, 71
company.....	(Used as example and not scored)					C, 2
competition.....	17	30	48	73	91	C, 66
consumer.....	14	36	72	87	89	D, 49
exploit.....	12	15	19	19	20	A, 70
export.....	41	75	93	94	98	C, 67
factory.....	(Not included in test)					
goods.....	46	83	97	99	99	A, 39
import.....	15	66	95	97	98	A, 45
industry.....	14	25	85	89	95	A, 40
manufacture.....	41	80	97	97	98	A, 38
merchandise.....	46	83	97	99	99	A, 39
production.....	18	34	53	77	79	D, 70

*This symbol is used to indicate those words the author would not now include in a list of essential terms, but upon which test data are available.

APPENDIX D

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—Continued

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
property.....	(Not included in test)					
raw material.....	40	42	72	80	90	A, 43
rebate.....	17	30	35	67	86	C, 64
shipping.....	(Not included in test)					
trade.....	18	52	84	95	99	B, 47
corporation.....	8	13	18	34	53	B, 22
monopoly.....	16	29	43	65	82	C, 68
trust.....	8	13	18	34	53	B, 22
employee.....	32	48	78	92	99	C, 44
employer.....	(Not included in test)					
labor.....	34	77	87	95	99	B, 10
strike.....	30	72	91	97	97	D, 25
union.....	32	48	78	92	99	C, 44
bank.....	(Not included in test)					
bankrupt.....	61	82	96	98	98	D, 23
bond.....	28	74	77	90	90	C, 31
capital.....	5	6	44	76	89	D, 29
credit.....	14	56	72	91	93	D, 21
crisis.....	24	33	59	80	88	C, 51
depreciation.....	28	46	74	82	88	A, 41
finance.....	41	73	85	90	95	B, 5
inflation.....	33	34	38	38	26	B, 53
investment.....	35	64	83	92	94	C, 72
market.....	48	72	85	93	97	C, 65
panic.....	19	26	39	49	77	A, 66
speculation.....	19	26	39	49	77	A, 66
stocks.....	54	81	95	96	97	B, 49
communication.....	24	57	93	96	97	D, 22
public utilities.....	18	34	45	62	73	D, 28
transportation.....	41	63	92	92	93	A, 67
prosperity.....	28	58	80	93	93	D, 47
wealth.....	49	84	84	85	86	B, 4
aristocrat.....	15	58	83	90	96	D, 15
peasant.....	32	73	91	93	97	B, 72
slave.....	56	67	95	95	97	A, 64
society.....	(Not included in test)					
community.....	45	68	94	98	98	C, 45
homestead.....	(Not counted because of error in score sheet)					
pioneer.....	24	69	84	89	93	C, 43

*This symbol is used to indicate those words the author would not now include in a list of essential terms, but upon which test data are available.

TESTS AND MEASUREMENTS

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—Continued

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
plantation.....	(Used as example and not scored)					C, 4
settlement.....	50	61	90	90	96	A, 69
rural.....	17	29	56	73	63	B, 70
urban.....	17	29	56	73	63	B, 70
census.....	41	63	89	95	96	D, 16
inhabitant.....	27	52	83	91	97	B, 17
population.....	27	68	94	95	98	B, 67
negro.....	28	71	84	85	88	D, 75
race.....	28	71	84	85	88	D, 75
mob.....	66	75	92	92	97	B, 74
riot.....	39	70	86	89	89	D, 64
emigration.....	11	65	86	93	97	C, 36
expansion.....	30	30	63	67	83	D, 60
immigration.....	33	48	89	96	97	A, 51
migration.....	29	42	72	87	96	D, 17
emancipation.....	22	25	59	69	86	B, 24
freedom.....	84	89	99	99	99	A, 33
independence.....	37	42	76	76	85	A, 63
liberty.....	84	89	99	99	99	A, 33
oppression.....	5	15	25	40	59	D, 67
education.....	(Not included in test)					
institution.....	(Not included in test)					
invention.....	39	52	89	90	96	A, 65
reform.....	19	44	53	65	85	D, 20
people.....	(Not included in test)					
private.....	(Not included in test)					
public.....	(Not included in test)					
public opinion.....	16	47	81	95	96	B, 48
standard of living.....	23	70	87	96	96	B, 33
*arbitrary.....	31	36	52	55	58	D, 42
illegal.....	32	33	66	80	96	B, 37
*invalid.....	6	11	13	24	24	B, 32
justice.....	39	52	80	87	92	A, 61
legal.....	(Not included in test)					
rights.....	25	54	76	85	88	B, 25
unconstitutional.....	6	11	13	24	24	B, 32

*This symbol is used to indicate those words the author would not now include in a list of essential terms, but upon which test data are available.

APPENDIX D

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—Continued

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
alien.....	20	39	88	97	99	A, 11
citizen.....	48	63	74	86	76	B, 39
exile.....	14	33	63	86	95	B, 36
nationality.....	48	67	90	93	96	C, 25
native.....	43	65	81	90	96	D, 10
naturalization.....	37	60	81	93	96	D, 26
appeal.....	(Not included in test)					
case.....	(Not included in test)					
convict.....	(Not included in test)					
crime.....	37	75	88	88	90	D, 77
decision.....	17	34	83	93	97	A, 57
execution.....	18	39	84	89	92	A, 72
injunction.....	7	19	42	55	69	B, 68
judge.....	(Used as an example and not scored)					A, 5
jury.....	29	74	84	89	95	B, 64
testimony.....	28	61	70	85	92	C, 73
verdict.....	17	34	83	93	97	A, 57
violation.....	21	53	93	94	94	D, 44
witness.....	28	61	70	85	92	C, 73
court.....	64	88	95	97	98	B, 52
jurisdiction.....	21	26	48	59	85	A, 74
supreme court.....	36	48	78	88	97	B, 38
bribery.....	20	34	70	88	96	A, 52
corruption.....	16	27	32	57	73	B, 34
fraud.....	31	56	63	89	94	D, 8
graft.....	16	27	32	57	73	B, 34
conspiracy.....	2	6	19	34	39	B, 75
insurrection.....	5	13	41	56	77	B, 71
rebellion.....	(Used as example and not scored)					C, 5
revolt.....	5	13	41	56	77	B, 71
revolution.....	(Used as example and not scored)					C, 5
secession.....	16	21	56	58	89	B, 13
sedition.....	2	6	19	34	39	B, 75
smuggle.....	37	52	80	87	92	A, 60
treason.....	30	48	76	93	94	C, 34
allies.....	34	37	73	82	91	B, 73
*belligerents.....	9	34	36	49	56	B, 23
enemy.....	14	38	64	76	92	B, 12
hostile.....	14	38	64	76	92	B, 12

*This symbol is used to indicate those words the author would not now include in a list of essential terms, but upon which test data are available.

TESTS AND MEASUREMENTS

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—Continued

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
*pirate.....	50	71	95	96	98	A, 55
army.....	(Not included in test)					
commander.....	(Not included in test)					
confederate.....	14	37	67	82	88	B, 45
cruiser.....	31	42	78	80	88	A, 78
general.....	(Not included in test)					
fleet.....	48	82	91	91	94	C, 79
forces.....	(Not included in test)					
naval.....	41	82	96	96	99	B, 60
navy.....	31	42	78	80	88	A, 78
officer.....	(Not included in test)					
reenforcements.....	(Not included in test)					
service.....	(Not included in test)					
troops.....	(Not included in test)					
submarine.....	76	97	97	99	99	D, 6
marine.....	15	39	52	58	72	D, 7
militia.....	2	12	22	38	44	A, 83
recruit.....	20	20	47	66	81	A, 35
soldier.....	(Not included in test)					
veteran.....	37	84	93	98	99	C, 21
volunteer.....	41	73	91	93	97	B, 63
draft.....	38	62	81	87	95	B, 3
mobilization.....	10	22	29	43	70	C, 55
*aggression.....	5	17	45	51	79	D, 9
attack.....	(Not used in test)					
battle.....	(Not included in test)					
blockade.....	43	43	69	72	83	B, 79
*bombardment.....	33	53	66	77	88	A, 56
campaign.....	19	30	50	59	81	D, 19
contraband.....	31	52	57	58	68	D, 50
defensive.....	41	80	90	94	97	C, 49
embargo.....	9	15	63	58	80	A, 81
*evacuation.....	12	16	24	37	39	B, 9
*fortification.....	33	53	66	77	88	A, 56
invasion.....	5	17	45	51	79	D, 9
*massacre.....	39	75	89	95	98	D, 46
military.....	34	69	81	86	95	C, 54
*munitions.....	18	34	56	59	77	D, 33
*occupation.....	12	16	24	37	39	B, 9
offensive.....	*1	80	90	94	97	C, 49
siege.....	6	23	59	73	81	A, 54

*This symbol is used to indicate those words the author would not now include in a list of essential terms, but upon which test data are available.

APPENDIX D

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—*Continued*

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
*strategic.....	13	22	53	61	77	D, 57
surrender.....	37	86	91	94	96	D, 14
victory.....	(Not included in test)					
war.....	(Not included in test)					
armistice.....	24	24	26	29	35	A, 62
disarmament.....	33	60	79	87	88	D, 56
indemnity.....	18	28	46	65	82	D, 32
reparations.....	23	23	65	74	91	B, 59
boundary.....	43	89	95	96	97	B, 77
continent.....	33	79	91	95	97	C, 24
*continental.....	4	10	11	13	20	D, 80
coast.....	(Used as example and not scored)					A, 2
frontier.....	(Item unsatisfactory)					B, 66
prairie.....	20	54	75	86	89	B, 54
district.....	(Not included in test)					
region.....	(Not included in test)					
section.....	(Not included in test)					
discover.....	53	71	90	93	97	C, 22
expedition.....	53	71	90	93	97	C, 22
exploration.....	(Not included in test)					
navigation.....	41	71	86	94	97	B, 6
voyage.....	59	97	98	99	99	D, 5
conservation.....	19	30	46	58	68	C, 23
natural resources.....	9	37	86	91	95	A, 19
agriculture.....	19	92	97	99	100	C, 7
irrigation.....	37	85	93	96	96	C, 6
reclamation.....	9	33	61	79	84	B, 7
*Pan-American.....	24	29	32	46	64	C, 74
clergy.....	22	27	50	76	86	B, 27
missionary.....	33	80	93	93	97	C, 70
creed.....	12	23	68	76	90	A, 82
*denomination.....	28	42	50	55	74	C, 17
*heresy.....	11	24	50	62	72	D, 68
intolerance.....	(Not included in test)					
persecution.....	11	24	50	62	72	D, 68
tolerance.....	26	27	41	58	76	A, 46

*This symbol is used to indicate those words the author would not now include in a list of essential terms, but upon which test data are available.

TESTS AND MEASUREMENTS

APPENDIX D: Percentage of Children in Each Grade Recognizing Each of the Words Tested—Continued

Words being Defined	Grades					Form and No. of Item
	4	6	8	10	12	
catholicism.....	26	26	58	63	67	A, 27
protestantism.....	26	26	58	63	67	A, 27
crusade.....	11	40	74	86	92	C, 10
papacy.....	23	30	52	79	88	B, 62
pope.....	23	30	52	79	88	B, 62
ancient.....	11	61	70	76	87	C, 8
century.....	25	80	92	94	98	B, 14
current.....	28	57	84	86	91	C, 14
decade.....	18	26	46	49	69	C, 13
era.....	25	41	44	72	83	C, 50
event.....	21	38	64	68	84	A, 76
mediaeval.....	19	32	47	72	81	D, 37
modern.....	11	61	70	76	87	C, 8
period.....	25	41	44	72	83	C, 50
history.....	(Not included in test)					
records.....	(Not included in test)					
propaganda.....	11	25	48	71	80	C, 33
publicity.....	(Not included in test)					
civilized.....	28	40	43	62	84	C, 63
primitive.....	28	40	43	62	84	C, 63
movement.....	(Not included in test)					
precedent.....	12	13	23	30	46	A, 22
tradition.....	12	13	23	30	46	A, 22

CLASSIFICATION OF GRAPHS

	Number of Each Type
1. Curves rising from low achievement in Grade IV to high achievement in Grade XII.....	70
2. Curves showing slow start, but rapid subsequent rise.....	31
3. Curves showing rapid initial rise, and then slow improvement.....	32
4. Curves showing high mastery in all grades; little increase.....	45
5. Curves showing very slow, but regular, progress.....	36
6. Curves containing a marked plateau.....	19
7. Curves showing a high start, but little subsequent improvement...	6
8. Curves showing marked gains between Grades VI and VIII, but little gain elsewhere.....	80
9. Curves showing almost no gains at any level.....	14
10. Curves in which a higher grade scores below the previous one....	13
Total.....	346

*This symbol is used to indicate those words the author would not now include in a list of essential content, but upon which test data are available.

THE WESLEY TESTS IN SOCIAL TERMS

EDGAR B. WESLEY
University of Minnesota

I. THE PROBLEM

The efficacy of objective tests in the measurement of information has been widely recognized. Their success in measuring understanding has been by no means so complete. Specific information which clusters around individuals, events, circumstances, and symbols has been measured quite successfully by the new-type tests, but even high scores do not prove that the testee *understands* the facts which he knows. Can objective tests be so constructed as to measure, or at least indicate, that the testee both knows and understands? The attempt to answer this question in the affirmative led to the construction of the Wesley Tests in Social Terms.¹

At the outset it was recognized that the usual array of names, dates, and events would test little beyond information. Specific facts, having no wide application, are likely to be practically inert, whereas terms expressive of social relationships have wide application and so are relatively functional. Consequently the first task was to select those terms which are functional, and the second task was to construct tests which would demonstrate whether the testee knew the term as a mere verbalism or whether he understood its true significance.

II. SELECTION OF TERMS

The word list from which the test terms were taken was compiled under the supervision of Professors Krey and Kelley in the fall of 1929. An effort was made to select those words and terms which have wide applicability. This involved the elimination of proper names and highly technical words and phrases. The bases of selection led naturally to the selection of words which function in more than one situation. Consequently the list was not limited to history or to any one of the social studies but was highly composite in its nature. When completed it contained about 4,000 words and terms.

¹*Wesley Test in Political Terms*, Forms A, B, C, and D; *Wesley Test in Social Terms*, Forms A and B; *Wesley College Test in Social Terms*, Forms A and B. New York, Scribners, 1932, 1933.

TESTS AND MEASUREMENTS

After it was completed it was sent to thirty-five college instructors, twenty-seven of whom cooperated in the evaluation of the words and terms. They were asked to indicate if the entering freshman should have (1) a fairly complete understanding of the word or term, (2) a partial knowledge, or (3) if no knowledge of the word or term was expected. The twenty-seven judgments on each term were combined into a single average rating which indicated the supposed utility and, by implication, something of the difficulty of each term. Thus the list of terms in the social studies represented the combined judgment of more than forty persons, including the working staff.

III. WESLEY TEST IN POLITICAL TERMS

The compilation of a word list and the evaluation of its contents in terms of utility and difficulty greatly facilitated the selection of items for the test. In addition to the official word list, considerable help was obtained from others which had already appeared in print.¹ For the purposes of the first test the author took the official (American Historical Association)² list and selected from it the *political* terms, taking care to omit (1) very simple words, (2) obsolete terms, (3) highly technical terms, (4) exceptionally difficult terms, and (5) phrases whose members could be utilized separately. Political was interpreted to include military, diplomatic, legal, and other terms closely related to government. To these words, numbering 1,190 were added 84 terms drawn from (1) Stephenson's list, (2) from a list of political terms on the junior high school level (prepared by Miss Edith Ware from the original A.H.A. list), and (3) words which suggested themselves to some member of the Commission staff as possessing great utility.

Words of varying degrees of difficulty were selected from this list of political terms to be utilized in the first edition of the Wesley Tests. Some terms were omitted because of the difficulty of reducing them to objective form. Special effort was made to utilize the

¹The two which were especially useful were Orlando Stephenson, "The Special Vocabulary of Civics," *Journal of Educational Research*, XVIII (Nov., 1928), 297-304, and E. E. Eubanks, "The Concepts of Sociology," *Social Forces*, V (March, 1927) 391-394.

²See appendix to volume.

WESLEY TEST IN POLITICAL TERMS

terms which appeared in all the word lists. A larger proportion of the terms in the important and difficult group was utilized than in the middle groups. Terms from local, state, national, and international fields were used so as to insure a representative sampling.

Edition I was prepared in the summer of 1930. It consisted of 274 items in a variety of forms: (1) historical illustration, (2) identification from a rather full description, (3) matching (three types within this general group), (4) substitution (both of the recall and control type), (5) completion (both of the recall and control type), and (6) best-answer. In September, 1930, Edition I was given to three hundred and ninety-seven freshmen at the University of Minnesota and to about two hundred seniors in high school. Under the direction of Professor Donald G. Paterson, a statistical study was made of the college papers and on the basis of these results items were selected for Edition II.

The results of Edition I were studied by means of both statistics and logical analysis. The high school papers as well as the college papers furnished material on which to base tentative conclusions. The process of evaluating the various forms was a prolonged task involving both statistical results and judgment. The true-false type was not used, for in the field of the social studies most truths are relative, and it is difficult to phrase an item involving either clear and unquestioned truth or palpable error. The form which called for an historical illustration proved to be a good measure of both information and understanding, but the form required the careful attention of one familiar with the field and, in the final analysis, rested upon judgment. In other words, it was not susceptible of mechanical objectivity. The identification from rather full descriptions required enormous space and was likewise devoid of complete objectivity. The matching forms seemed to allow too large an element of guessing, especially after the first few operations had been performed. The same objections may be made to the substitution types. The completion type works well when concerned with specific facts, but is much more difficult to construct when dealing with functional ideas and is lacking in objectivity in its answers. On the other hand, the best-answer type with five op-

TESTS AND MEASUREMENTS

tions was selected as the form which called for the highest degree of discrimination. The testee, being aware of the fact that the correct answer is before him, needs to employ rote memory scarcely at all. Many items contain acceptable options but only one which is clearly best. The form allows the use of false, but attractive, associations, of partially correct ideas, and of a variety of options based upon similarities of sound and form. The student who picks his way through such an array of misleading suggestions and insists upon choosing the best answer has, allowing for the small percentage of chance, demonstrated his understanding more convincingly perhaps than by any other type of objective form. The director of the Commission and most of the staff agreed that the best-answer type showed most promise, and it was accordingly accepted as the one to be used most extensively.

The items which survived Edition I were utilized in Edition II. No new items were introduced, but some were changed from other forms to the best-answer type. Two parallel forms of fifty items each were prepared and given to 1,132 high school and college students. Edition II was given to the following numbers in the various grades:

<i>Grade</i>	<i>Numbers</i>
11.....	252
12.....	376
13.....	409
14.....	26
15.....	39
16.....	30

The returns from Edition II were studied under the direction of Professor Truman L. Kelley. Forty of the hundred items seemed to have genuine promise. They were selected on the basis of the statistical results and were published in four forms of ten items each. Each form has a reliability coefficient of .68, which is sufficiently high for the measurement of groups, but not sufficiently high for the measurement of individuals. When all four forms are used, however, the resulting score is sufficiently reliable for individual diagnosis. The test is standardized for the 12th grade, but may

WESLEY TEST IN SOCIAL TERMS

be used with the three grades before and after. Its usefulness extends from the 9th grade to the end of the sophomore year in college.

The validity of a test is the degree of success with which it measures what it purports to measure. Validity, in the final analysis, must rest upon judgment, although statistical returns will either confirm or tend to invalidate judgment. The validity of the *Wesley Test in Political Terms* rests upon the following: (1) The material utilized in the tests was taken from official and approved sources. (2) The material unquestionably belongs within the field of government. (3) Only those items which differentiated reliably between the superior and inferior groups were retained in the final forms. (4) The correlations between the *Wesley Test in Political Terms* and other tests (reported fully in the Examiner's Manual accompanying the test) show that it measures to some extent what intelligence tests measure, and to a somewhat greater degree it measures what reading and civic tests measure. But the correlations, even when corrected for attenuation, fall far short of perfect relationships. The conclusion is that the Wesley Test measures to a considerable degree what other tests do not measure. The assumption is that it measures understanding in the social studies.

IV. WESLEY TEST IN SOCIAL TERMS

The same lists used in the construction of the test in *Political Terms* were used in constructing the test in *Social Terms*. The lists by Eubanks and Stephenson and indexes of texts in sociology, economics, and political science were consulted in the determination of doubtful cases. This test differs fundamentally from the first in the respect that it utilizes terms from all the social studies instead of being confined to political history and government.

Several of the items used in the trial forms of the test in *Political Terms* were utilized in what might be called Edition IV. All of the conclusions as to form and technique were so utilized.

The resulting study showed that 160 items possessed a satisfactory power to discriminate. They were accordingly arranged in two parallel forms of 80 items each. All that was said about the

TESTS AND MEASUREMENTS

validity of the test in *Political Terms*, save the correlation with other tests, holds true in regard to the test in *Social Terms*.

V. WESLEY COLLEGE TEST IN SOCIAL TERMS

In addition to the A.H.A. list and college texts, another source of first importance was used in the construction of the *College Test*. In the summer of 1932 the author had charge of students who enrolled for a symposium on the Social Studies which was held at the College of Education at the University of Minnesota. The lectures were combined into a series to constitute a course for which college and graduate credit was given. Two rather elaborate tests were constructed in connection with the course. A careful evaluation of the items was made by Mr. Wilbur Murra of the University High School. Thus, some sixty items of recognized validity were available as a nucleus for the college test. The college test utilizes terms from all the social studies. It is, in fact, a more advanced form of the test of political terms which was first evolved for the twelfth grade. The college test is standardized for the end of the sophomore year in college, but may be used for all years from the 12th grade to the senior year in college.

Another factor of undoubted significance was the prolonged, critical, and patient help of members of the social science departments of the University of Minnesota. They added materially to the validity of the test by their careful evaluation of each item.

The preliminary forms, consisting of 137 items in each of two forms, were given to students in four widely scattered colleges as follows:

<i>Class</i>	<i>Numbers</i>
Sophomores.....	44
Juniors... ..	101
Seniors.....	87
Total.....	<u>232</u>

A statistical study of the results was carried on under the direction of Dr. F. H. Finch, Head of the Personnel Department of the University of Minnesota High School. From this study, 210 items of recognized power to discriminate were arranged in two parallel forms of 105 items each.

CONCLUSIONS

What was said as to the validity of the test in *Political Terms*, save the correlation with other tests, applies equally well to the college test.

·VI. CONCLUSIONS

The varying connotations of words is a familiar experience to adults. Even the simplest words enshroud themselves in habiliments which strike persons other than the user as strange or even outlandish. When one has passed beyond the simplest and most elementary uses of a word, it begins to assume these varying connotations. The most careful student of language is likely to be most keenly aware of the various meanings and complications of a word or term. The person who has grasped the word as little more than a verbalism is likely to be the one who thinks he knows its full and correct meaning. As one advances in experience and learning, he realizes the endless stages or levels of meaning and so becomes less dogmatic as to final connotations.

New words, upon their initial appearance, seem to be more or less mechanical or arbitrary names, which are for the sole purpose of identifying the thing or idea to which they are applied. In this initial stage they serve the purpose in a manner as devoid of personalities as if they were numbers. A toy horse might just as well be called by another name so far as the child is immediately concerned. *Constitution* is an arbitrary word for the pupils of the middle grades until it becomes more than a mere identification tag. *Imperialism* is a highly abstract idea, perhaps only a sound or a series of letters, to the high school boy who encounters it for the first time. In each instance the word may remain as a mere name or it may become the symbol of a rich and many-colored idea.

But even verbalisms have their utility and constitute the first step in the acquisition of an understanding of a word. As the pupil moves through the successive grades he acquires both fuller connotations of familiar words and introductions to new ones; so in high school and college, as well as in the elementary grades, the teacher is confronted with the problem of the learning process through all of its stages. It is probable that the high school stu-

TESTS AND MEASUREMENTS

dent has begun to grasp the higher levels of some words and so is prepared to recognize the possibilities of enriched connotations. These advances, however, are by no means regular, nor do they necessarily correlate highly with either grade classification or intelligence. The measurement of these levels of understanding as opposed to mere verbalisms is a task that requires the constant attention of the teacher of the social studies.¹

An ideal test would measure not only the highest but the intermediate levels of understanding. Each item would be so graduated as to reveal the level which the testee has reached. It would also take account of verbalisms, since they probably indicate the first step on the road toward a full understanding. Many tests do provide for this measurement of verbalisms or the recall of rote information; fewer of them call for the discrimination which is involved in an understanding of ideas; and no test in the social studies provides for graduated scoring within the various options of the items. The *Wesley Tests* do not systematically allow for this graduated degree of correctness within each item; in this respect they would seem to fall short of what is possible in dealing with some terms. The extent to which this graduated scheme can be applied is yet undetermined.

In spite of the limitless nature of words, they do have, at various levels, fairly definite meanings, the understanding of which can be tested. The *Wesley Tests* attempt to measure this understanding. They are designed to indicate the degree of comprehension or level of understanding which a student has of a particular word. The simplest, most obvious meaning is not always that which is sought; frequently a meaning further toward the periphery is the one involved. The care with which the tests were made, the rigid criticism through which they passed, and the thorough revisions which were made, together with the preliminary reports upon their use, seem to indicate that they have to some considerable extent achieved the purposes for which they were designed.

¹A fuller discussion of a parallel idea may be found in Edgar B. Wesley, "Facts or Ideas in the Social Studies," *Historical Outlook*, XXIV, 28-30; A. C. Krey and Edgar B. Wesley, "Does the New-Type Test Measure Results of Instruction in the Social Studies," *Ibid.*, XXIII, 7-21.

THE KELTY-MOORE TEST OF CONCEPTS IN THE SOCIAL STUDIES

MARY G. KELTY AND NELLE E. MOORE

I. INADEQUACY OF CHILDREN'S CONCEPTS

Language and thought have long been recognized as interdependent. The purpose of language is to express thought; and thought, on any except the simplest plane, is impossible without language. "The relation between thought and language becomes so intimate that the two are psychologically inseparable."

Yet it is fallacious to assume that children understand an idea because they can repeat the words in which the idea is embodied. Those words may not, in any way, link themselves up with the children's background of experience, and may therefore completely lack meaning, even for those who use them glibly.

Many investigations have concerned themselves with children's concepts. G. Stanley Hall in his study, "The Contents of Children's Minds,"¹ in 1890 showed that children habitually used words which had little or no meaning for them. Earl Barnes² investigated specific terms in history and found that children understood poorly all the terms except those which touched their own experience directly. Chambers³ went a step farther and classified the causes of children's misapprehensions of terms, such causes, for example, as similarities in sound or form. Scott and Myers⁴ suggested, as a result of their investigation, that children's errors expressed language difficulties in which "the whole *modus operandi* of thought is involved" and that "much of school learning is only

¹G. Stanley Hall, *Some Aspects of Child Life and Education*, Ginn & Co., 1907 Ed., 1-52.

²Earl Barnes, *Studies in Education*, Vol. II. Stanford University, California, 1897. p. 60.

³Will G. Chambers, "How Words Get Meaning," *Pedagogical Seminary*, XI (1904) 30-50.

⁴Flora Scott and Garry C. Myers, "Children's Empty and Erroneous Conceptions of the Commonplace," *Journal of Educational Research*, VIII (1923) 327-34.

word manipulation." Meltzer¹ tested the concepts of children in succeeding grades, beginning with the fifth, to determine both their specificity and their completeness. He found that growth was steady throughout the grades, but that misapprehension and vagueness were common. Ayer² showed that the progress of many children in studying history was blocked by their lack of comprehension of abstract thoughts and words, as well as by "literary embellishments" and involved sentences. Piaget³ pointed out in his reports another factor of significance, *i. e.*, that nine- to eleven-year-old children not only did not understand many words and sentences which they used fluently, but that they believed they did understand the ideas and asked for few explanations.

After considering such investigations, one is inclined to agree with John Dewey that vague meanings are "the source from which flow most bad intellectual consequences." To determine whether or not children possess clear-cut and vivid concepts of the meanings connected with words becomes, therefore, a problem of vital concern to the teacher. Measuring or testing concepts is a means of measuring or testing thought.

II. LISTS OF THE MOST IMPORTANT SOCIAL CONCEPTS

In order to be able to test concepts, the teacher must understand the general processes by which concepts are developed. A discussion of these processes, however, lies outside the field of concern of the present article.

If, in addition to understanding the development of concepts, the teacher is to be able to ascertain whether or not children have a real understanding of terms essential to the social studies, she should possess a guidance list of the most important social concepts. The assembling of such a list ought not to be left to the individual judgment of each teacher. It ought at least to represent a pooling of the judgments of experts.

¹Herman Meltzer, *Children's Social Concepts*. Bureau of Publications, Teachers College, Columbia University, 1925.

²Adelaide M. Ayer, *Some Difficulties in Elementary School History*. Bureau of Publications, Teachers College, Columbia University, 1926.

³Jean Piaget, *Language and Thought of the Child*. Harcourt Brace and Co., 1926.

THE KELTY-MOORE TEST: ITS CONSTRUCTION AND VALIDITY

Such a list has been prepared for publication as a part of the report of the Commission on the Social Studies of the American Historical Association. Numerous lists have also appeared for the separate subjects, and there are even some lists for special subjects at a particular school level.¹

For the purpose of the present investigation (the Kelty-Moore Test of Social Concepts), the above-mentioned list of the Commission on the Social Studies was used as a basis. To this list were added at the lower level many simpler terms and concepts, and many terms from Industrial History, from the Kelty list referred to in footnote 1, p. 233.

III. THE KELTY-MOORE TEST: ITS CONSTRUCTION AND VALIDITY

In order to test the understanding of concepts on the part of children in the middle grades, as well as in the higher grades, the simplest possible form was chosen.² (See pages 9 ff.) The procedure was as follows:

(1) The composite list of words issued by the Commission on the Social Studies (referred to in the second paragraph above) was first divided into groups of terms which, in the opinion of the testers, could be satisfactorily taught to children in the following grades: fourth, fifth, sixth, and junior high school. The simpler concepts and those used primarily in the early periods of American History were assigned to the fourth grade; those from later American History, to the fifth grade. Since the most common subject of history study in the sixth grade is European Background, the terms used chiefly in that field were assigned to the sixth grade. More difficult terms in American History, including complex political and economic concepts, were assigned to the junior high school.

It is obvious that many fourth grade to junior high school pupils

¹The lists for history have been summarized in *The Journal of Educational Research*, XXXIV (December, 1931), 335-49. For the A. H. A. list see appendix to this volume.

²Rather than such forms as used by Garr and Gifford in the *Second Yearbook of the National Council for the Social Studies* (1932), p. 164; or by Shaffer in the *Journal of Educational Psychology*, XIX (1928), 41-44.

TESTS AND MEASUREMENTS

do not know the specific terms assigned to their grades because they have never been exposed to the materials. The testers, however, are convinced from years of experience in teaching those particular terms and from informal testing following the teaching, that average children of the grades mentioned can understand the terms if they are taught them directly.

(2) Each of the graded lists of terms described above was then divided into groups according to meaning. The purpose of such grouping was to assist in random sampling and to insure the inclusion of all the different types of words. The scheme of classification according to meaning was:

Objects with which children have had experience.

Objects with which children have not had experience.

Names of actions.

Mental attitudes.

Condition or state of being.

Political and economic concepts.

Miscellaneous (words which did not belong in any of the other categories).

(3) Test-lists of fifty-six terms each were then selected for all the grades under consideration, by choosing eight terms from each of the seven categories. The total number of words in each category was divided by eight, and the count proceeded on the quotient thus secured.

(4) The list of fifty-six terms for each grade was divided into two approximately equivalent forms, X and Y.

(5) With the assistance of the Standard Dictionary and Webster's Dictionary a correct statement of the meaning of each term was written. An effort was made to avoid the exact dictionary definition. A second statement of meaning was also prepared, which was nearly right, or which, in the judgment of the testers, represented errors that children would be likely to make.

(6) A list of the fifty-six terms for each grade (fourth to junior high school) was then given to one group of children from each grade with the request that they write a statement of what they thought each term meant. From these responses, enough wrong

PROGRESS IN LEARNING

items were selected to make a total of five items for each vocabulary-term.

(7) Forms X and Y for each grade from the fourth to the junior high school were then mimeographed. Each test was administered to two groups of children in each grade (about fifty children in most cases).

(8) The results were tabulated. The tabulations of certain of the items called the attention of the testers to misleading statements and to other unwise complications in the test. Accordingly, changes were made where necessary, to insure that the test should measure only what it was designed to measure, and that the concept itself should present the only difficulty.

(9) The completed test-forms were then presented to three leading authorities on the teaching of the social studies, who criticized them on the basis (1) the importance of the terms, and (2) the adequacy of the statements from considerations of scholarship.

(10) These tests were then administered to a random sample of 100 fourth grade pupils, a random sample of 100 sixth grade pupils, and a random sample of 100 eighth grade pupils. The items were split into halves as follows: (1) items 15-36, 46, 48-51, 66-71, 79; (2) items 1-13, 37-42, 44, 53-54, 56-60, 63, 72, 74-77, and correlated to determine their reliability. The poorer items which had been rejected during the process of construction were also split into halves and correlated as follows: (1) items 45, 47, 52, 64-65, 80, and (2) items 14, 43, 55, 61-62, 73, 78. The correlations resulting are tabulated herewith.

RELIABILITY OF ORIGINAL KELTY-MOORE TEST ITEMS

	Better Items	Equivalent Reliability per Item	Poorer Items	Equivalent Reliability
100 4th grade pupils.....	.749	.082	.028	.0044
100 6th grade pupils.....	.765	.089	.290	.059
100 8th grade pupils.....	.790	.101	.329	.070

Seventy items were thereupon chosen. These were divided into two forms (A and B) of thirty-five items each for individual placement, the reliability per form being in the neighborhood of .77.

TESTS AND MEASUREMENTS

All seventy items in one test furnish a reliability of .85 in the fourth grade, .87 in the sixth, and .90 in the eighth.

In addition seven approximately equivalent forms (Forms 1-7) of ten items each were arranged for survey purposes. The reliability per grade group was about .45.

(11) The eighty items from which were later chosen the seventy items which constituted Forms A and B, and Forms 1, 2, 3, 4, 5, 6, 7 were administered to pupils in grades four to nine in schools at Dearborne, Michigan; Montevallo, Alabama; Reading, Massachusetts; Washington, D. C.; Wilmette, Illinois; and Winnetka, Illinois. The tentative norms upon Forms A and B, and Forms 1, 2, 3, 4, 5, 6, 7 were computed from this tryout. They appear below.

TENTATIVE NORMS ON THE KELTY-MOORE TEST OF CONCEPTS IN THE SOCIAL STUDIES

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
Form A or B.....	9.0	13.0	18.6	23.4	26.0	26.2
Forms 1, 2, 3, 4, 5, 6, 7....	2.6	3.7	5.3	6.7	7.4	7.5
No. of Pupils.....	216	199	170	260	316	181

IV. THE USES OF THE TEST

Forms A and B and Forms 1-7 serve to give a general picture of children's understanding of essential terms in the social studies, chosen by random sampling. They assist teachers in determining how a given group compares, in general, with other groups throughout the country. They may point out the necessity of revising the local program of instruction, in such respects as wider use of extensive reading, more inclusion of actual experiences, more direct instruction in word-meanings, and more systematic testing of the understanding of terms.

If the problem of testing children's concepts is to be attacked constructively and continuously, however, the Kelty-Moore Test will be serviceable only as a guide and a measuring stick. The real solution of the problem lies in the breaking-up of the complete list of social studies concepts into small groups, whose meanings can be

THE USES OF THE TEST

developed in connection with a small unit of experience.¹ This short list should be made the subject of direct instruction, such as the supplying of experiences or the recalling of experiences. After a proper interval, the short list should also be made the subject of a test similar to the Kelty-Moore Test.² Where necessary, the instruction of individual pupils should be repeated. Only by such a systematic, continuous program can teachers be assured that children possess real concepts rather than a list of words empty of meaning.

¹For example, as in Mary G. Kelty, *Teaching American History in the Middle Grades of the Elementary School*. Ginn and Co., 1928, pp. 56, 61, 64, 68, 72, 76, 80, etc.

²May be secured through Charles Scribner's Sons.

CHAPTER IV

TESTING OF SKILLS

The objective of skills was described by the planning committee as follows:

"Skill in the use of sources of information about society

A student who has taken courses in history and other social studies is expected to know how to find and use intelligently the various sources of information relating to current society.

Such courses include:

Current gossip—oral and printed.

Reasoned discussion—oral and printed.

Social activities—real or pictured.

Material achievement—in use or relic.

The skills necessary to deal with such sources range from simple physical findings and manipulation to subtle critical analysis. A series of tests ranging from the simple to the more complex phases of these skills must be constructed. Some are already available."

H. E. Wilson of Harvard University collaborated with the staff of the Investigation in trying to formulate a check list of "skills" or "working habits" as those were recognized by teachers. One of his students, Price, conducted a questionnaire inquiry of teachers, the tabular results of which are here set forth.

It was early apparent that so little work had been done in the measurement of such skills as the social studies required that only pioneer work could be attempted by this Investigation. The work of Parker and Calkins, directed primarily at the testing of geographical knowledge and understanding, promised to blaze a wide trail into the field of graphic skills. Marion Clark and E. M. Hunt undertook to test the less graphic skills associated with critical reading. Clark succeeded in devising some elementary tests of historical criticism which are here described.

REPORT OF THE WORK OF THE COMMITTEE ON GEOGRAPHY TESTS¹

EDITH PUTNAM PARKER
University of Chicago

INITIAL DISCUSSION AND PROPOSALS

In 1930, members of a proposed geography committee were invited to confer with representatives of the Commission for the Investigation of the Social Studies in the Schools about the advisability and feasibility of testing, under the auspices of the Commission, for fundamental geographic attainments believed to be of value in the pursuit of social studies. The advisability of so doing hinged in part on the relation of geography to the social studies. Modern geography is a science which may be termed "socio-natural" to distinguish it from the social studies and from the natural sciences. If the testing work of the Commission was to be restricted to the purely social studies, tests in geography obviously could not be included legitimately. All participants in the discussion recognized, however, that the progress of pupils in the social studies probably is facilitated by the possession of various geographic understandings, abilities, and attitudes; they thought, accordingly, that data of value to the Commission and not available from other sources could be secured by the use of geography tests focused on such attainments. The geographers realized that the sponsorship of the proposed tests by a Commission on Social Studies might give the false impression that geography was being included *as a social study* rather than *as a science that contributes, among other things, to intelligent, effective pursuit of social studies*. On the other hand, they appreciated and welcomed the opportunity to be of service, and realized that experimentation with such tests should be of value also in improving the teaching of geography. In view of the latter considerations, it was decided to risk the possibility of misimpression noted, and to undertake the task, if, upon other grounds, the project seemed desirable and feasible.

Decision with regard to the desirability of the undertaking

¹The tests are published by Charles Scribner's Sons.

TESTS AND MEASUREMENTS

hinged on the question, "May the tests be concerned wholly with *basic values inherent in sound geographic training?*" It is recognized widely that, in spite of recent improvement in the teaching of geography in the schools, actual procedure in many geography classes centers attention largely on the memorization of miscellaneous facts. It fails to lead pupils to derive, from carefully guided use of relevant facts, many of the fundamental values of geographic training. Obviously, then, testing for evidences that those basic values had been derived would not be synonymous with testing for what had been accomplished in the schools under the flag of geography. It would show only what had been accomplished in the way of fundamental geographic outcomes.

The geographers involved made it clear that they had no interest in measuring progress along lines *not* directed toward basic geographic goals. They expressed the conviction that tests designed for measuring so-called geographic progress, instead of basic geographic progress, not only yield findings of little or no value, but also tend materially to retard improvement in the teaching of geography, and in the services that it may render to the social studies. They called attention to the general belief that an achievement test indicates what experts think should be achieved, and to the fact that teachers accordingly tend to set up standards of achievement in terms of the abilities and understandings which their children need in order to make a reasonable showing on achievement tests. They submitted that improvement in geographic training can be brought about through tests only to the extent to which testing emphasis is placed on fundamental geographic achievements.

Viewing the matter from a positive rather than a negative angle, they claimed that tests stressing geographic essentials would serve to increase an understanding on the part of the educational public, of values that are inherent in geography, and would tend to stimulate and guide efforts to derive those values.

The project, then, seemed *undesirable* unless the tests could stress achievements resulting from the most valuable type of geographic training.

INITIAL DISCUSSION AND PROPOSALS

The question of feasibility also hinged largely on this issue of testing for "what ought to be" rather than for "what is." Even if it were desirable to discover what is being done, regardless of its value, such a task would have been impracticable in view of the necessary restrictions as to length and number of the tests. The facts that comprise most courses in "geography" are exceedingly numerous and varied. General testing experts long have been baffled by the problem of formulating brief, comprehensive tests in cases where facts are multitudinous and of many types. Pupils in class A, with as much fact baggage as those in class B, or even more, may make a much lower score than the latter on a given "fact test" because the particular facts chosen were not in their repertoire. With another sampling, class A might seem superior to class B. It is clear that *sound* conclusions concerning what is being achieved in "fact courses" can be based only on many and varied samplings of each representative group of pupils tested.

The difficulty disappears in large measure if tests center on major understandings, abilities, and attitudes to be derived from the *use* of relevant facts, for the fundamental attainments of these types are not so numerous as to preclude effective sampling in relatively brief tests. In such a case, one tests for many facts, as the later discussion will show, but he tests for them not as ends in themselves but as elements in abilities and understandings. He tests for power to use facts effectively in the types of situations in which their use is helpful. For example, a test which involves *reading* facts of various types from a *given* physical map puts emphasis on a functioning ability. It samples adequately one's ability to read other maps of similar kind, and the elements in map-reading ability are few enough so that each can be touched upon in a relatively brief test exercise. A pupil needs to know many facts in order to read such a map well. Evidence of ability to read the map thus is evidence also of possession of these facts. Evidence of lack of this ability, or of particular phases of it, is evidence also of (1) lack of the necessary information concerning basic facts, or (2) inability to make that information function in the practical situation in-

TESTS AND MEASUREMENTS

volved. The same general points may be made in connection with any type of major geographic understanding or ability.

As a result of the discussion along the foregoing lines, the representatives of the Commission proposed:

- (1) that it be made clear in the report of the Commission that geography was involved not as a social study, but as one which has contributions to make to the pursuit of social studies;
- (2) that a geography committee construct experimental tests of *achievements which represent basic geographic attainments*;
- (3) that the tests be not limited to what is achieved generally, but that they be based on what it is practicable to achieve through the best type of geographic training which can be given; and
- (4) that the tests be objective, comprehensive, group tests planned for use with college freshmen.

To understand the reasons for the nature of the tests constructed and for the handling of the data secured through their use, it is essential to bear in mind the special type of test which was specified in this agreement.

Mr. R D Calkins, of the Department of Geography, State Teachers College, Mount Pleasant, Michigan, and Miss Edith Parker, of the University of Chicago, were asked to cooperate with Dr. Bowman in the undertaking. The Commission felt that the experiences of these three geographers would enable them to make diversified contributions to the work of the committee.

PRELIMINARY WORK AND PLANS

In planning the procedure to be followed in fulfilling the assignment, the committee recognized the necessity of (1) reviewing studies already made concerning the fundamental values to be achieved through geographic training, (2) comparing the findings of those studies with the beliefs of professional geographers as expressed or evidenced in texts and in other published or unpublished material, (3) outlining the types of understandings and abilities

PRELIMINARY WORK AND PLANS

which, according to the results of (1) and (2) probably should be tested for, and (4) securing the reactions of others to that outline.

A series of studies, later reported in the 1933 Yearbook of the National Society for the Study of Education, had shed much light on the problem of selecting major types of achievement which are distinctively geographic. The technique used in those investigations had precluded the possibility of favoring any particular delimitation or definition of the subject. *All* types of materials called "geographic" by their authors had been analyzed, and, by the comparison of the types of ideas and abilities developed thereby with types of ideas and abilities developed in materials *not* labelled geography, those distinctive in geography had been ascertained. There was some disagreement among members of the geography committee itself as to the proper scope of the subject and as to the terms that best describe some of the distinctive types of ideas developed therein, but these differences were of little or no significance in the task at hand. To define geography, to settle differences of opinion concerning its terminology, and to attempt a delimitation of its scope were not among the assigned duties. It was obvious from the reviews of studies, of texts, and of opinions of experts that there existed a body of fundamental, distinctive understandings and a group of abilities needed in acquiring them effectively, upon which essentially all geographers seemed agreed, regardless of their definitions of the subject, and of their favorite terms for describing these types of outcomes. The committee accordingly adopted the policy of centering attention on those abilities and understandings upon which all seemed to be in accord. The ability to read maps for geographic purposes, for example, was so conspicuously a part of every proposed program of geographic training, was so emphasized in every text in the subject, and was so clearly envisaged by every exposition or philosophical discussion of geography that it seemed altogether safe to list various phases of the ability to read maps among those which would be sanctioned by all geographers as representing one of the basic values of the study. The major types of understandings and abilities listed were:

TESTS AND MEASUREMENTS

UNDERSTANDINGS

1. Of the interrelations of cultural and natural features which characterize specific *regions*, such as France, western United States, a given desert region in low latitude, and the like.
2. Of the nature, conduct, and world distribution of given kinds of human activities (*e. g.*, the manufacture of rubber goods) as related, on the one hand, to the cultural assets of the people engaged in them and, on the other hand, to the natural factors these people capitalized or coped with in carrying them on.
3. Of various ways in which peoples of different cultural assets capitalize or cope with specific types of natural factors, such as climate, topography, and the like.

ABILITIES

The ability:

1. To read from landscapes, pictured or real, facts of significance in acquiring these understandings.
2. To read such facts from maps.
3. To read such facts from graphs and statistical material.
4. To acquire such facts from the printed page (verbal reading matter).
5. To reason accurately in terms of such facts.
6. To express effectively geographic information and geographic conclusions based on such reasoning.
7. To apply to new situations understandings and abilities of the types indicated.

Practically all authors of texts and of courses of study in geography for elementary schools, state, in essence, that one objective of the work is to give understandings of the first general type named. Understandings of the second type are stated, in effect, to be a goal of work outlined in texts and courses for secondary schools. Such courses commonly are designated by the terms "commercial," "economic," or "political" geography. In courses of study commonly called "principles of geography," "elements of geography," "physical geography," and the like, and in texts used in them, the giving of the third general type of understanding customarily is evidenced as an objective. In connection with the

PRELIMINARY WORK AND PLANS

gaining of any of these types of understandings, texts and courses of study stress the desirability of developing abilities of the various types noted. In so far as their expressed purposes are concerned, authors of these guide materials all seem to seek these fundamental achievements. Obviously the general emphasis in current school-room practice on the mere memorization of facts is not in harmony with the avowed purpose of experts in the field. It seemed clear from the expressions of purpose examined that the list represented types of achievements that at least the vast majority of geographers would agree upon as of fundamental importance.

Dr. Bowman, who had conducted a separate survey of opinions for a different purpose, felt, on the basis of his findings, that this was the case. To secure a further check, the committee, in conference with representatives of the Commission, listed geographers at fifteen centers of geographic training as those to whom the list should be sent first for criticisms. It was felt that the group selected represented a fair sampling of expert geographers and that, if they offered no objections to the list, the committee could be reasonably certain that it was a sound one on which to proceed. In the letter accompanying the list, the geographers addressed were asked to express freely their opinion concerning the advisability of testing for attainments of the types indicated rather than for mere facts as ends in themselves. Every reply sanctioned the former procedure and condemned the latter. No objections were raised as to the types of attainments listed. Two queries were made about terms used in describing the understandings, but it was made clear in each case that there was no objection to the essence of the thing expressed.

As a result of this inquiry, the committee felt reasonably sure that to base a test on the understandings and abilities listed would meet with the general approval of professional geographers. It is significant, as bearing out that conclusion, that as the tests themselves were presented later, all of the many geographic critics who reacted to them seemed to accept the attainments involved as unquestionably of basic value.

It next was decided to make a test consisting of several "parts".

TESTS AND MEASUREMENTS

and to devote one part to each of the three major types of understandings listed. Part I, for example, was to be devoted to testing for understandings of the first type named. It was obvious, however, that it would be impracticable to devote a separate part to each type of ability indicated. Experimentation showed that graphs and pictures could be used in testing for some of the understandings. Thus, a person who connected with a specific region facts shown by a group of graphs would evidence not only a knowledge of that region but also an ability to read significant items from those graphs. It was realized that an incorrect reaction to such an exercise might be due to lack of knowledge of the region, to lack of ability to read the graphs accurately, or to both. Some light would be thrown on the nature of the difficulty, however, by comparing this reaction with others. Map reading, unlike landscape and graph reading, involves so many phases of ability that it was impossible to test adequately for ability in it in connection with testing for understandings. In view of this fact and of the importance of developing these various phases of the ability, it was decided to devote a separate part of the test to map reading. As one basis for the diagnosis of failures evidenced in connection with the other four parts, it was planned to devote a fifth part to a vocabulary test dealing with technical and semi-technical terms. The general plan of the test may be summarized as follows:

Part I. World understanding in terms of understandings of regions commonly introduced in elementary geography: some of the exercises to be based on pictures and graphs.

Part II. Map reading ability: exercises to be based on accompanying map book.

Part III. World understanding in terms of understandings of activities commonly treated in secondary school geography: some of the exercises to be based on pictures and graphs.

Part IV. World understanding in terms of understandings developed in high school courses of the types noted in connection with the third major type of understandings: some of the exercises to be based on pictures and graphs.

Part V. Vocabulary test.

THE CONSTRUCTION OF THE TEST

THE CONSTRUCTION OF THE TEST

It was decided to construct first the part of the test dealing with map reading. This procedure would make it possible to carry forward work on the map books which were to accompany the exercises while other parts of the test were being prepared. Moreover, it seemed probable that experience with this part, which could be completed with less difficulty than any other, would facilitate the construction of the other parts. Selection of items was relatively simple because it hinged upon such definitely ascertainable facts as (1) the types of maps for which one has need in everyday affairs, (2) the types of symbols used on such maps, and (3) the types of ideas which are expressed in terms of these symbols. Examination of current texts, newspapers, periodicals, atlases, and maps indicated that the types of maps used most are (1) black and white "sketch" maps, consisting of symbols for such features as cities, railroads, rivers, and mountains; (2) automobile road maps; (3) political maps; (4) physical, or color-contour maps; (5) maps depicting such items as forest areas, wheat and other crop acreage, and rainfall; and (6) weather maps, topographic maps, and the like. The symbols used on these maps are of four major types, namely: (1) semi-pictorial signs; (2) color bands, color spots, and dots; (3) orientation lines—the network of parallels and meridians; and (4) "iso" lines.

Semi-pictorial symbols are those that suggest by their form the features they symbolize. The symbol for a coast line, for example, suggests the appearance of an actual coast line as seen from the air. Practically all symbols on sketch maps and automobile road maps, many of those on political and physical maps, and some of those on nearly all maps are of this type. Because they partake somewhat of the nature of a picture or drawing, they are the simplest type of symbol to read. In the accurate interpretation of symbols of the second type named, careful use of accompanying legends is required because these symbols are not semi-pictorial. Dots may stand, for example, for a given acreage of wheat on one map, for a designated production of rye on another, and for a given number

TESTS AND MEASUREMENTS

of sheep, or for something entirely different, on a third. Accordingly, the interpretation of these symbols is less simple than that of symbols of the semi-pictorial type. The colors on political maps and on color-contour maps, and the colors, shades, or dots on various distribution maps are of this second type. Orientation lines may be used on any type of map, and they are found almost invariably on the political and physical maps in common use. The translation of line symbols of both the orientation type and the iso-line type into the conditions they depict demands a higher order of ability than that required for the interpretation of symbols of the first two types indicated.

The kinds of ideas to be read from semi-pictorial symbols are those of (1) the *nature* of the features symbolized (including their size, extent, and shape, as well as their general type) and (2) the *relative location* of particular features with regard to other features (direction and comparative distance). The chief function of color bands, color spots, and dots is to give ideas concerning the *areal distribution* of the features, conditions, or activities represented. The purpose of orientation lines is to give (1) ideas of *distance and direction from the equator and prime meridian*, and (2) ideas of the *relation of any part of the earth's surface to other parts of its surface and to its surface as a whole*, in terms of size and location. Iso-lines convey (1) ideas of *exact elevation, slope, air pressure, wind direction, temperatures*, and the like, and (2) ideas of the *areal distribution* of the conditions they depict.

In making use of the foregoing groups of facts, the committee based its work on the following considerations:

1. However valuable a map may be for use in everyday affairs, it is useful to a given individual only to the extent to which *he* can get from it the ideas it is designed to convey.

2. It follows that testing for map-reading ability of a type which is desirable to develop in the course of elementary and secondary education means testing for the ability to get from maps, of types that function helpfully in everyday affairs, the ideas those maps express.

THE CONSTRUCTION OF THE TEST

3. In accordance with the initial agreement concerning the nature of the test, this part would be based on map-reading ability needed in everyday affairs and *not* on a consideration of that possessed by most graduates of secondary schools.

4. Findings from the use of such a test probably would show that many, or even most, of the persons tested possessed only a few phases of this ability. Such a result would not point, however, to the impracticability or undesirability of developing the other phases, but merely to the fact that the experiences of those tested had not been such as to insure that development. Until one knows what phases of map-reading ability have resulted from the experiences pupils have had, and what phases of it have not resulted, one has no sound basis for providing more helpful experiences. Only a test constructed on the basis of what constitutes desirable map-reading ability, then, would be of value in a program of improvement.

In planning the map book to accompany the exercises, it obviously was desirable to provide, in so far as practicable, for maps of each of the several types above listed. The expense of colored maps was prohibitive, but it proved feasible to include all other types.

The necessity for brevity led to the construction of exercises of a true-false type, and to the inclusion in many cases of only one exercise for a given phase of ability. It is to be noted, however, that none of the true-false statements was such as to call for mere memory. All of them were based on the accompanying maps, and one can judge of their truth or falsity only in terms of what he *reads* from the maps to which they refer.

As a basis for checking the relative effectiveness of these true-false exercises, multiple choice exercises dealing with the same phases of ability were devised. Two forms of true-false and two forms of multiple choice tests were constructed. In each form, exercises were so grouped as to facilitate the use of the accompanying map books. The arrangement was such that the turning of pages, for example, was minimized. Grouping also was such as to accord with standard chance distribution theorems. In the true-

TESTS AND MEASUREMENTS

false forms, approximately fifty per cent of the exercises were followed by those calling for reactions of the opposite type; twenty-five per cent were followed by two exercises calling for reactions of the opposite type; twelve and a half per cent, by three opposites; six and a fourth per cent, by four opposites, and six and a fourth per cent, by five or more opposites. The number of "trues" equalled, of course, the number of "falses." In the multiple choice forms, care was taken to group the statements from which selections were to be made so that the first statement was correct in approximately twenty-five per cent of the cases, the second in approximately twenty-five per cent of the cases, and so on. They also were arranged to avoid any set order: an exercise in which the first statement was correct was followed in some cases by one in which the first statement also was correct, and in other cases by one in which the second, third or fourth statement was correct.

In the course of the work on the map test, preliminary experimentation was carried on with a group of about two hundred and seventy students in the Teachers College at Mount Pleasant, Michigan. The chief purpose was to ascertain the amount of time needed to react to various types of exercises. The conclusion was reached that the test must be limited to about fifty exercises, to meet the time requirements set by the Commission. It was decided to include exactly fifty items, each to be given one score point, in the belief that this procedure would minimize work with scores.

Before these tests, designed wholly for college freshmen, were completed, the committee was told of the intention of the Commission to use the tests also in all grades from the fourth to the twelfth inclusive. This intention did not introduce any new factor into the work save that it led to a reconsideration of the language used with a view to making it as simple as possible. All the other considerations on which the work was based held for the testing of pupils throughout the grades indicated. Any test which would serve to show what phases of map-reading ability college freshmen possessed and what phases they did not possess would serve to show the same things in the case of pupils in any lower group who could handle the terms used.

TESTING THE ABILITY TO READ MAPS

The following table indicates the specific phases of map-reading ability covered, and the numbers of the questions in each form which were designed to test the respective phases.

PART II

TESTING THE ABILITY TO READ MAPS

SECTION A. USING SEMI-PICTORIAL MAP SYMBOLS

Specific Abilities Tested	Exercises Used for Testing			
	Form I	Form II	Form I	Form II
	T. F.	T. F.	M. C.	M. C.
Ability to use semi-pictorial map symbols to identify:				
sea and land.....	1	2	1	1
railroad.....	2	2
river.....	..	1
strait.....	3	..	3	..
mountainous land.....	2	..
cities and mountainous land.....	4
cities and coast.....	..	3
cities and rivers.....	4
cities and lowland.....	4	..
island.....	3
peninsula.....	..	4
Ability to get from semi-pictorial symbols ideas of:				
general direction.....	5	5	5	5
direction of stream flow.....	6	6	6	6
relation between rail route and river valley.....	7	8	7	7
comparative distance (without use of scale of miles).....	8	7	8	8
Ability to use scale of miles in determining distance.....	9	9	9	9
Ability to use ideas of relations between street patterns and natural conditions in interpreting facts shown about street patterns.....	10	10	10	10
Ability to use symbols and legend on conventional automobile road map to determine distance, direction, surfacing, and highway marking.....	11 & 12	11 & 12	11 & 12	11 & 12

TESTS AND MEASUREMENTS

SECTION B. USING COLOR BAND, COLOR SPOT, AND DOT SYMBOLS

Specific Abilities Tested	Exercises Used for Testing			
	Form I T. F.	Form II T. F.	Form I M. C.	Form II M. C.
Ability to use such symbols in reading facts concerning:				
elevation.....	13	13	13	13
density of population.....	..	13	..	13
Ability to use such symbols in comparing facts about:				
density of population and amount of rainfall.....	14	14
elevations and rainfall.....	..	14	14	..
Ability to see probable relationships between facts read about:				
production of crops and amount of rainfall.....	15	15
trade, production of crops, and density of population.....	..	15	15	..
topography, wind directions, and rainfall.....	16	16
temperature and altitude.....	..	16	16	..

SECTION C. USING ORIENTATION LINES

(The Network of Parallels and Meridians)

Specific Abilities Tested	Exercises Used for Testing			
	Form I T. F.	Form II T. F.	Form I M. C.	Form II M. C.
Ability to use meridians and parallels in reading directions from a network in which meridians and parallels are straight lines:				
north-south.....	17	17	17	17
east-west.....	..	17	..	17
Ability to use meridians and parallels in reading directions from a network on which meridians and parallels are curved lines:				
east-west.....	18	..	18	..
north-south.....	19	18	..	18
northeast-southwest.....	20	19 & 20
Ability to read from a straight-line network facts about:				
direction from equator.....	21	..	19	..
latitude in terms of degrees.....	22	21	20	20

TESTING THE ABILITY TO READ MAPS

SECTION C. USING ORIENTATION LINES—Continued

Specific Abilities Tested	Exercises Used for Testing			
	Form I T. F.	Form II T. F.	Form I M. C.	Form II M. C.
latitude in terms of high, low or middle.....	24	19
latitude in terms of comparison with distance from equator to pole.....	..	22	..	22
degrees of latitude as translated into miles.....	..	24	22	..
Ability to use facts thus read about latitude in explaining:				
type of vegetation.....	23	21
length of growing season.....	..	23	21	..
length of days.....	30
Ability to read from a curved-line network:				
latitude in terms of degrees.....	25	27
latitude in terms of high, low or middle.....	..	26	33	..
latitude in terms of comparison with distance from equator to pole.....	27	..	24	..
degrees of latitude as translated into miles.....	28	26
Ability to use facts thus read about latitude in explaining:				
type of crop grown.....	..	25	25	..
length of growing season.....	26	25
length of day.....	..	28	26	..
Ability to read from a straight-line network facts about:				
direction from prime meridian.....	29
longitude in terms of degrees.....	30	..	28	..
Ability to use facts thus read in:				
estimating distance in miles.....	23
comparing time.....	29
Ability to read from a curved-line network facts about:				
direction from prime meridian.....	..	29	27	..
longitude in terms of degrees.....	..	30	..	24, 28
longitude in terms of comparison with distance from prime meridian to 180th meridian.....	31	27

TESTS AND MEASUREMENTS

SECTION C. USING ORIENTATION LINES—*Continued*

(The Network of Parallels and Meridians)

Exercises Used for Testing

Specific Abilities Tested	Exercises Used for Testing			
	Form I T. F.	Form II T. F.	Form I M. C.	Form II M. C.
Ability to use facts thus read in:				
estimating distance.....	..	31	23, 29	..
comparing time.....	32	32	30	..
Ability to compare extent of an area shown on straight-line network with extent of one shown on curved-line network.....	33	33	31	31
Ability to identify map of given region on one network with same region on another network.....	34	34	32	32
Ability to explain seeming differences in the shape of a given region on two maps	33
Ability to use straight-line network in reading:				
difference in latitude of two given places.....	35	35	35	35
difference in longitude of two given places.....	36	36	34	..
difference in size of areas extending through the same number of degrees of longitude and latitude, but in different latitudes.....	38	38	36	34
Ability to identify given latitude and longitude with specific continents and oceans.....	37	37	35	36
Ability to:				
read directions from a polar projection.....	39	39	37	37, 38
identify meridians and parallels on a polar projection.....	40	40	38	39

SECTION D. USING ISO-LINES OF VARIOUS TYPES

(ISOTHERMS, ISOBARS, AND CONTOURS)

Exercises Used for Testing

Specific Abilities Tested	Exercises Used for Testing			
	Form I T. F.	Form II T. F.	Form I M. C.	Form II M. C.
Ability to use isothermal maps in reading and comparing facts about:				
average monthly temperatures at two or more places in a given month.....	41	41	39	..

TESTING THE ABILITY TO READ MAPS

SECTION D. USING ISO-LINES OF VARIOUS TYPES

(ISOTHERMS, ISOBARS, AND CONTOURS)

Exercises Used in Testing

Specific Abilities Tested	Exercises Used in Testing			
	Form I T. F.	Form II T. F.	Form I M. C.	Form II M. C.
range in temperature between two sets of places.....	42	42	..	40
average temperature for two or more given months.....	43	43	41	41
average temperature in a given month and latitude.....	40	..
range in temperature and location..	42
Ability to use facts thus read in explaining crop productions.....	44	44	42	..
Ability to use isobaric maps in reading facts about:				
pressure and wind direction.....	45	45	43	44
pressure and wind velocity.....	43
Ability to use facts thus read in interpreting weather conditions.....	46, 47	46, 47	44, 45, 46	45, 46
Ability to use contour maps in:				
reading facts about elevation.....	48	48	47	49
identifying topographic profiles.....	50	50	50	..
identifying topographic features....	49	..
Ability to use facts thus read in helping to explain:				
choice of routes.....	49	..	48	..
selection of town sites.....	..	49
flood hazards.....	50
Ability to identify isobars, isotherms, isohyets, and contours.....	47, 48

Though Part II was the least difficult to construct, much time was consumed by the experimentation at Mount Pleasant, the analysis of its results from various angles, the construction of the maps for the map book, and the reworking of the tests to meet (1) the various conditions outlined above and (2) minor defects in wording revealed by critical study of the exercises and informal experimentation with them. In the four months the committee could devote wholly to the work, the map tests were completed, Part I was planned in detail, and the remaining parts were blocked

TESTS AND MEASUREMENTS

out roughly. It was impossible after that time for any two members of the committee to work in the same place, and the necessity of conducting the work by means of correspondence slowed it materially. Part I was completed in this way, but little more was done with the last three parts. General plans made for them will be discussed in a later section of the report.

The procedure in preparing Part I of the test was similar to that described in detail for Part II. The purpose, as noted earlier, was to test for an understanding of the world in terms of a knowledge of outstanding characteristics of its various regions. Since some of the characteristics of any region are associated closely with its latitude, several exercises were devoted to testing for the ability to read from landscapes, descriptions, and graphs, features or facts indicative of latitude; and for the ability correctly to associate those features or facts with such location. For example, "The places shown in Figures 1 and 2 [pictures] are probably about the same distance from the equator." Correct reaction involves a knowledge of the types of vegetation typical of low latitudes and of middle latitudes, and an ability to recognize those types in terms of the broad-leafed (banana, etc.) trees shown in one picture, and the grass lands shown in the other.

"The region [described in the preceding paragraph] is near the south pole." In that paragraph, facts are stated which are true only of places in *high south* latitude. Knowledge of conditions near the south pole enables one to react correctly.

"Since June is one of the cooler months there, the country is south of the equator." The statement refers to a temperature graph. If one reads the graph correctly he finds the statement false because it shows June is one of the warmer months in the region in question. Ability to read the graph, and knowledge of the time of occurrence of seasons north of the equator and south of it enable one to react correctly.

The exercises in Part I which deal with this general ability are listed below. The second column indicates the type of the source from which facts were to be read, and the third column the types

TESTING THE ABILITY TO READ MAPS

of association or relationship, knowledge of which is needed to make correct reactions.

Group I. Recognizing signs of location with regard to the equator and associating them properly with such location.

Item	Source of significant facts	Types of associations involved	
1	picture	nature of vegetation.....	relative distance from equator
3	picture	nature of vegetation.....	conditions in northern and southern hemispheres.
4	picture	nature of vegetation.....	conditions in northern and southern hemispheres
22	description	continuous daylight in December.....	distance and direction from equator
52	description	nature of crops.....	distance from equator
59	description	nature of crops and growing season.....	direction from equator
60	description	nature of crops.....	distance from equator
61	description	nature of crops and growing season.....	extent in latitude
76	graph	time of occurrence of seasons.	direction from equator

About three-fourths of Part I tested for the ability to identify with specific regions given groups of facts or relationships which are outstanding characteristics of the regions in question. Great stress was put on this phase of world understanding for the following reasons:

1. It is the alleged purpose of elementary instruction in geography to give pupils correct impressions of the modes of life in various regions of the world, and to help them properly to associate human activities characterizing these regions with natural conditions typical there.

2. If one *has* learned the geographic characteristics of a country, he should have no difficulty in identifying them when they are depicted in landscapes, in descriptions, or in graphic representation.

3. Of the various achievements that result from sound geographic training, none is more useful in contributing to progress in the social studies than a knowledge of the geographic personalities of various regions of the world.

TESTS AND MEASUREMENTS

Care was taken to select regional characteristics of importance and to deal only with regions or places frequently referred to in current literature and treated in practically all elementary texts.

The test items devoted to this ability to identify the chief geographic characteristics of various countries or places were of two kinds. One of them may be thought of as constituting Group II and the other as constituting Group III. In items in Group II, the characteristics to be identified were *shown in pictures, described verbally or depicted in graphic form*. The persons tested were asked to tell whether the facts thus presented fit given regions or places, or do not. In items in Group III, the matching was in reverse order. A given region or place was named, and various facts or relationships listed. In the case of each fact or relationship those tested were asked to tell whether or not it helps to characterize the region in question. In items of the Group III type, several reactions are called for in the case of some of the countries or places to be characterized. The time to which administration of the test to a given group had to be limited made it impossible to treat many regions in this way.

Examples of exercises of the Group II type follow.

"There are places much like the one shown in Figure 1 [a picture] in central Africa."

"This description fits many places in the Sahara and southwestern Asia."

"All facts shown [in the accompanying graphs] are true of Egypt."

The following lists indicate the items in Group II, the types of sources from which characteristics were to be read in each case, the types of characteristics to be associated with the regions in question, and the regions with which they were to be identified.

Group II. Identifying given facts or groups of facts or relationships with specific regions or places.

Source of significant facts	Types of associations involved
10 picture nature of	vegetation southernmost North America

TESTING THE ABILITY TO READ MAPS

Item	Source of significant facts	Types of associations involved
11	picture..... nature of vegetation.....	southernmost South America and Australia
12	picture.. nature of vegetation..	northernmost Africa
13	picture.. nature of vegetation..	central Africa
14	picture.. nature of vegetation..	southern Europe
15	picture.. nature of vegetation..	northern South America
16	picture... vegetation, surface, and human activity...	the Nile region
17	picture.. vegetation, surface, and human activity...	the Congo region
18	picture.. vegetation, surface, and human activity...	the Amazon region
19	picture.. vegetation, surface, and human activity..	Switzerland
20	picture.. vegetation, surface, and human activity	Norway
21	picture.... vegetation, surface, and human activity...	Holland
25	description... nature of houses, food, and work..	Sahara and southwestern Asia
26	description.. nature of houses, food, and work....	southwestern United States
30	description... nature and distribution of work, crops, and vegetation	northern United States and southern United States
31	description... nature and distribution of work, crops, and vegetation	eastern United States and western United States
32	description... nature and distribution of work, crops, and vegetation	Virginia, Kentucky, Missouri, and Kansas
33	description... nature of work and location in given part of the United States...	Lowell [Massachusetts]
34	description... nature of work and location in given part of the United States.....	Atlanta [Georgia]
35	description... nature of work and location in given part of the United States.....	Birmingham [Alabama]
36	description... nature of work and location in given part of the United States.....	Pittsburgh [Pennsylvania]
38	description... nature of farms, of farm crops, and of farm work.....	New England

TESTS AND MEASUREMENTS

Item	Source of significant facts	Types of associations involved
39	description... nature of farms, of farm crops, and of farm work. .	States bordering Great Lakes on the south
40	description. nature of farms, of farm crops, and of farm work	States at eastern base of the Rocky Mountains
41	description nature of vegetation and temperatures	northwestern United States
42	Description... nature of vegetation and amount and distribution of rainfall .. .	northwestern United States
44	description .. kinds and distribution of work	Washington [state]
45	description... kinds and distribution of work	Oregon
46	description.. kinds and distribution of work	North Carolina
53	description.. kinds and distribution of work, density and movement of population	Chile
54	description... kinds and distribution of work, density and movements of population	Italy
55	description... kinds and distribution of work, density and movements of population	Spain
56	description distribution of population and railroads; important products and exports	Russia
57	description.. distribution of population and railroads, important products and exports	Canada
58	description . distribution of population and railroads, important products and exports	Australia
62	description .. surface, climate, density and distribution of population, standard of living, mineral resources, nature and distribution of work.	China
63	description... surface, climate, density and distribution of population, standard of living, mineral resources, nature and distribution of work.....	India

TESTING THE ABILITY TO READ MAPS

Item	Source of significant facts	Types of associations involved
64	description... surface, climate, density and distribution of population, standard of living, mineral resources, nature and distribution of work.....	Brazil
78	graphs..... climate, land utilization, crop acreage, imports and exports..... bar line pie	Union of South Africa
79	graphs..... climate, land utilization, crop acreage, imports and exports..... bar line pie	Egypt
84	description... location; nature of tributary area.....	Buenos Aires
85	description... location; nature of tributary area.....	Calcutta
86	description... location; nature of tributary area.....	Marseilles
87	description... location; nature of tributary area.....	Cairo
88	description... location; nature of tributary area.....	Sydney

Examples of exercises of the Group III type and list of items included in that group follow:

(These statements were to be evaluated as true or false.)

"Chicago, in part because it lies in a vast agricultural lowland at a convenient meeting place for lake and land routes, is the chief railroad center of the United States."

"The fact that Mexico is not a leading world power is explained in part by scarcity of mineral resources there."

"A major factor in the growth of London has been its location on one of the large producing coal fields of England."

Group III. Associating correctly with given places, countries, or groups of countries specific facts or relationships.

Item	Place or region given	Type of fact or relationship to be evaluated concerning it
47	northeastern United States	type of surface and transportation net
48	Chicago.....	location; nature of tributary area

TESTS AND MEASUREMENTS

Item	Place or region given	Type of fact or relationship to be evaluated concerning it
49	chief mining state in the United States	location; features in northeastern United States related to mining in this state
50	New Jersey . .	soil, drainage, density of population
65	Mexico	world power in relation to mineral resources
66	Mexico	world power in relation to farm, forest, and grazing resources
67	Mexico	world power in relation to climate
68	Mexico	world power in relation to human characteristics
69	Brazil	distribution of people
70	Brazil	mineral resources; forest resources; importance of agriculture
71	Brazil	surface and soil of eastern and southern parts
72	Brazil	outside communication in relation to nature of coast
73	Brazil	nature and distribution of agriculture in relation to soil and climate
80	London	growth as related to location and nature of river
81	London	growth as related to location with regard to coal fields
82	London	growth as related to natural route focus and government work
83	London	growth as related to location with regard to sea routes
89	southwestern Asia	sparsity of population as related to location with regard to sea routes
90	southwestern Asia	sparsity of population as related to location with regard to land routes
91	southwestern Asia	sparsity of population as related to climate
92	countries of Europe	colonizing activities
93	countries of North America	colonizing activities
94	countries of South America	political status (independence vs. dependency)
95	countries of Africa	political status
96	colonizing countries	colonizing activities in relation to size of colonizer
97	colonizing countries	colonizing activities in relation to density of population

TESTING THE ABILITY TO READ MAPS

Item	Place or region given	Type of fact or relationship to be evaluated concerning it
98	colonizing countries. . . .	colonizing activities in relation to latitude
99	colonizing countries. . . .	colonizing activities in relation to home resources in proportion to size of country
100	colonizing countries. . . .	colonizing activities in relation to home advantages for trade and manufacturing

A fourth group of items tests for the ability to associate correctly with given facts (read from pictures, descriptions, or graphs) *other* facts, *not* given, which are typically true of places where the given conditions are found. For example:

"People who live in places like that shown in Figure 1 [picture] prepare for long winters."

"In the state referred to in paragraph VIII [a description] canning factories, saw mills, and flour mills are probably more numerous than iron, steel, and textile mills." This involves an understanding of relationships existing between these types of manufacturing establishments (not mentioned in the description) and the production of specific types of raw materials mentioned in the paragraph.

"Population probably is evenly distributed throughout the area [represented by accompanying graphs]." This involves the ability to read from the graphs that rainfall in the region represented is scant, that more than ninety-five per cent of the area is desert, but that thousands of acres are cultivated; it involves also the ability to associate with those facts uneven distribution of population, about which the graphs gave no direct information.

The items included in Group IV are listed below, together with the types of sources from which the given facts were to be gleaned and the types of facts (not given in these sources) to be associated with those given.

TESTS AND MEASUREMENTS

Group IV. Associating correctly with given facts various other facts and relationships normally related to them.

Item	Source of facts given	Types of associations involved
2	picture..... use of materials for food, clothing and shelter.....	vegetation indicative of low latitude; of middle latitude
5	picture..... length of seasons.....	vegetation indicative of low latitude
6	picture..... amount of rainfall.....	vegetation indicative of low latitude
7	picture..... length of days.....	vegetation indicative of low latitude
8	picture..... position of noon sun.....	vegetation indicative of low latitude
9	picture..... length of seasons.....	vegetation indicative of low latitude; of middle latitude
23	description... amount of rainfall; irrigation.	nature of houses, food, work
24	description... day and night temperatures; length of seasons.....	nature of houses, food, work
27	description... aridity.	nature of houses, food, work
28	description... length of growing season.....	nature of crops grown
29	description... time of occurrence of season.	harvesting activities
37	description... length of growing season; amount of summer rainfall.	nature of crops grown
43	description... kinds of manufacturing work.	agricultural, mining, and logging activities
51	description... seasonal distribution of rainfall.....	nature of crops grown
74	graph..... distribution of population...	climate; land utilization
77	graph..... importance of agricultural work.....	nature of trade

One of the hundred items in the test cannot be classified as belonging to any of the groups designated. It is item 75: "Rainfall amounts to about 10 inches a year." It refers to a rainfall graph and tests purely for the ability to read the graph. It was included for diagnostic purposes.

ANALYSIS OF RESULTS

It should be clear from this analysis of the test that much fact knowledge was called for; it should also be clear that the test called for much *thought*. In a given description, *some* of the facts may fit a certain region, but not all of them. One must weigh the group carefully to reach a correct decision. To react correctly and consistently to many of the items, one must have knowledge of interrelated human and natural facts which characterize the regions or places in question and must be able to see relationships between outstanding types of facts.

It seems clear to the committee that unless knowledge and abilities of the types represented in Parts I and II do result from the study of geography in elementary and secondary schools, much of the opportunity of contributing to an ability to pursue social studies effectively is being lost.

ANALYSIS OF RESULTS

The administration of the tests and the compilation of data secured through their use were not under the auspices of the committee. The results were reported to the committee in the form of item profiles. In the case of each item in the tests there was recorded, for each grade or level in which that exercise was tried, the percentage of the pupils tested who reacted correctly. The college freshman year is designated as grade 13, and the Stanford and Princeton freshman groups were assigned to Grade 13+. These data were examined with a view to discovering what light they shed on the answer to five major questions, namely:

1. What evidence do the data furnish concerning the reliability or dependability of the exercises as instruments for taking an inventory of the stock of fundamental geographic achievements of groups of pupils?
2. In so far as the exercises are dependable, what do the results show about the attainments of the groups tested?
3. What do the results reveal concerning progress at various levels toward the goals involved?
4. What light do they shed upon weaknesses in the training of pupils in geography and upon remedial planning?

TESTS AND MEASUREMENTS

5. What light do they shed upon needs to be met in planning for further testing?

In answering these questions, results of the use of Part II were analyzed first. Consideration of the dependability of the items involved was based on the following two understandings:

1. A test item is dependable if it has validity, that is, if it tests for the specific ability for which it was designed to test, and if it has reliability, that is, if reactions to it are consistent with the ability known to be possessed by those who react to it.

2. Reliability of an item can be established only in terms, then, of *two* factors, namely: (1) reactions of individuals or groups to that item and (2) *other evidences* of possession of the ability involved, or of lack of it, on the part of the individuals or groups reacting.

All the data furnished to the committee had to do with the reactions of individuals and groups to items in the test itself. Though four forms of the map-reading test had been supplied by the committee for the purpose of securing several evidences concerning the abilities involved, *these forms were not administered to the same groups*. Reliability obviously cannot be established merely in terms of the performance of different groups with corresponding items. The data furnished, then, are inadequate for the purpose of determining dependability. However, certain general facts about the abilities of the groups tested are known, and the nature of the test itself is such that these data, *when considered in the light of what is known about the abilities of the performers*, throw some light on the matter.

As to validity, the nature of these questions is such that it is difficult to conceive of their testing anything other than the specific ability they are designed to test. The language is exceedingly simple and the statements are perfectly definite. Take, for example, the statement, "The latitude of J in Figure 9 is 60° N." If one can read latitude accurately, he can tell from the map that this statement is true and can react correctly. If he does not have that ability, he may react correctly by chance. If he reacts correctly to ~~this kind of statement repeatedly~~, or if a high percentage of a large group reacts correctly, the chance element is outweighed. What

ANALYSIS OF RESULTS

ability could consistently correct performance in this case possibly evidence other than the ability to read latitude from maps of the type used? Logic prompts one to answer "None."

"In Figure 12, arrow R is pointing due south." What ability other than the ability to read directions from a map drawn, as is "Figure 12," on a polar projection possibly could be evidenced by consistently correct reactions to this statement? What could consistently incorrect reactions evidence except inability to read directions from such a map?

Such considerations, though subjective, should have weight in evaluating testing material.

To gain further evidence concerning validity, the following experiment was conducted. A group of twenty-five adult students (college seniors and graduates) who were thought to possess the map-reading abilities involved were given several informal tests to secure evidence to that effect. When satisfactory evidence was in hand, these students were given the tests involved in this discussion and were asked not only to react to each statement, but also to state the specific ability for which each item tested. Twenty-one of them made perfect scores (50), and four made scores of 48 or 49, having erred either in their count of contour lines in determining elevation or in calculating range in longitude or in both. In no case did they fail to recognize the specific ability for which each given exercise called.

These exercises were given afterward to a group of fourteen students whose map-reading ability was an unknown quantity. Each individual of the group was tested separately on every phase of ability about which his test performance left doubt. It was established by this procedure that six per cent of the test errors were due to carelessness, and that 94 per cent of them were due to lack of requisite ability. These students, like the others, listed correctly the abilities for which the test called, even in the cases upon which they failed. They recognized what each exercise called for, whether or not they could respond correctly. The results of this experimentation supported the subjective conclusion concerning validity.

TESTS AND MEASUREMENTS

In the construction of the map test, the committee had provided carefully for another means of checking on dependability. The ability to recognize and name semi-pictorial symbols is one which practically every normal individual acquires relatively early in his experience with maps and one which he tends to retain. One recognizes some of these symbols almost as readily as he recognizes features in picture, because of their semi-pictorial character. He may recognize some of them without any technical training. To recognize certain combinations of these symbols which represent straits, peninsulas, and the like, he needs to learn what straits are, what peninsulas are, and so on. These terms are introduced relatively early in most, if not all, courses in elementary geography. This phase of ability constituted, in essence then, a known quantity. Care was taken to devote about a tenth of the test to items of this type calling for abilities which one scarcely could fail to have developed, whatever his geographic experiences had been. Consistently correct reactions to these items would show that test results relating to them were in accord with known facts concerning the ability of the groups tested. These items in Form I of Part II were 1, 2, 3, 4, and 5. The performance of all groups on these items was high. The lowest percentage of correct reactions by any group of pupils to any one of them was 84, which was the percentage of correct reactions that sixth grade pupils made to item three. This percentage was high for sixth grade pupils, for it was reached or exceeded in the case of only seven items out of the fifty involved. The lowest percentage in the college freshmen group was 94. Results from the use of this group of items, then, were in harmony with what were essentially "known abilities" on the part of those tested.

Some definite knowledge also is possessed concerning the ability of adults to use map networks effectively in determining the relative size of two areas. Dr. Helen M. Strong of the United States Department of Commerce experienced so much difficulty from this source on the part of business men seeking information that she brought about the publication by the department of an equal area world map for their use. It was designed to help in overcoming the

ANALYSIS OF RESULTS

inability to recognize distortion of areas on the mercator or any cylindrical projection. Item 38 tested for this ability to recognize the difference in size of areas (mapped on a cylindrical projection) that extended through the same number of degrees of latitude and longitude but were in different latitudes. Consistently low performance on this item would be in accord with knowledge concerning probable lack of ability of those tested. The percentage of correct responses on the item varied from 0 to 33. To no other item was there so low a percentage of correct responses.

Mr. F. E. Lord, director of the training of rural teachers at the State Teachers College, Ypsilanti, Michigan, had carried on a scientific investigation of the ability of seventh and eighth grade children to read latitude and longitude. His findings afford clear indication of the fact that pupils who possess these phases of ability are fewer than those who possess ability to read semi-pictorial symbols. In the light of his investigation, the percentages of correct reactions to items 22, 23, 25, 26, and 30 should be less than those to items 1, 2, 3, 4, and 5. The percentages of correct reactions by the sixth grade pupils to the questions ranged from 50 on item 30 to 69 on item 26, as against 84 or more on the semi-pictorial items. The percentages of correct reactions by college freshmen to these items ranged from 73, on each of items 22, 23, and 25, to 83 on item 30. The performance of these groups is typical of the performance of all groups tested. The results, then, are in harmony with those of Mr. Lord's investigation.

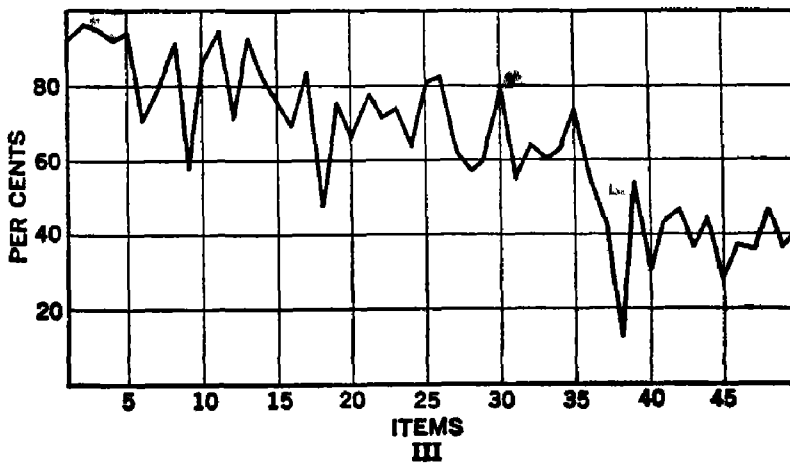
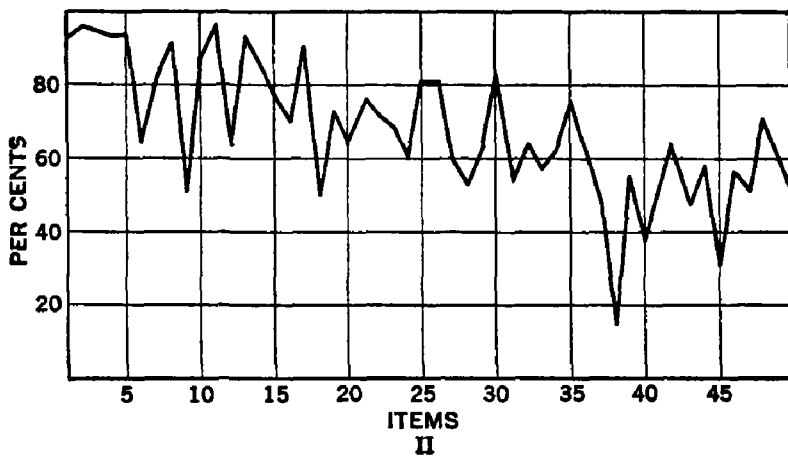
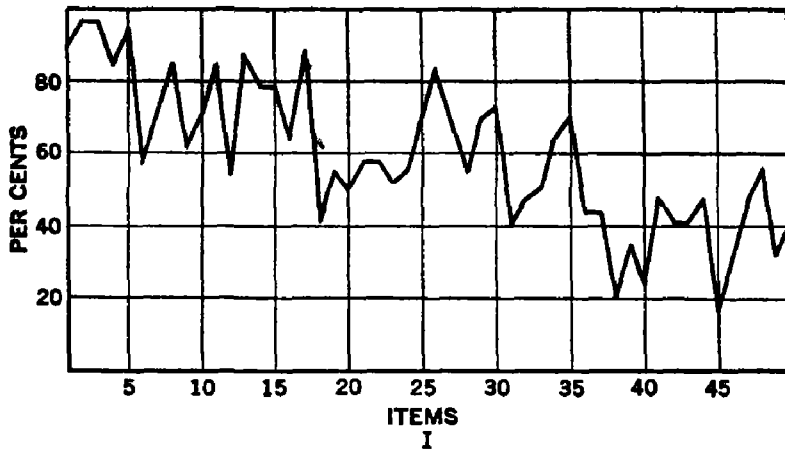
Other examples of similar character might be added, but these suffice to illustrate the point that *in so far as the knowledge of the second factor had been obtained, performance on the tests was in harmony with what was known about the possession of the abilities for which the exercises test.* Since all the test items were similar in form, were expressed in language as simple as the requisite technical terms permitted, and were different only in the types of ability for which they called, it seems altogether probable that (with exceptions to be noted later) all the items were as dependable as those for which data were available for comparisons of the type just indicated.

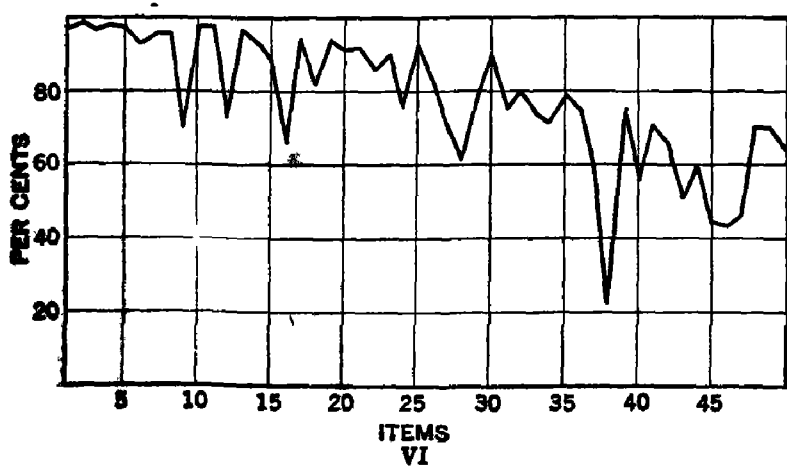
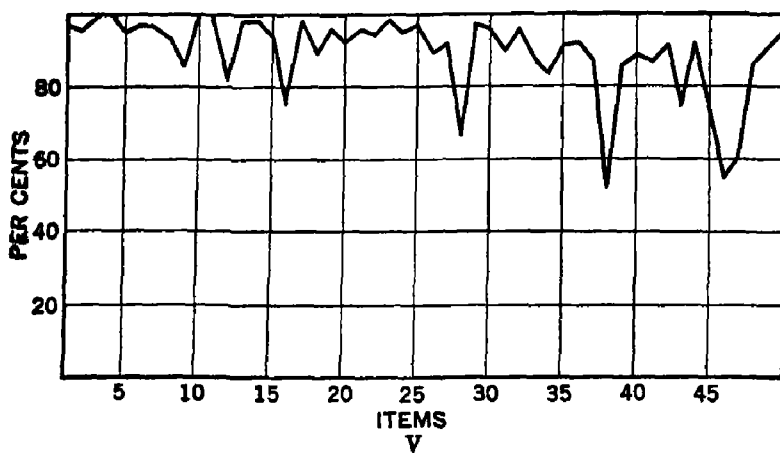
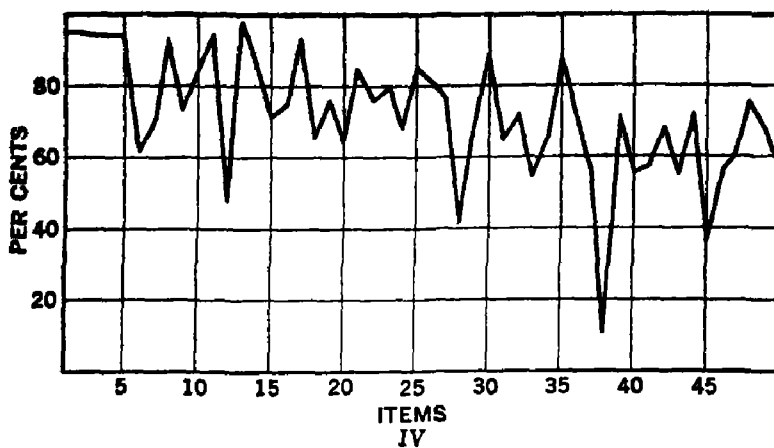
TESTS AND MEASUREMENTS

Comparison of the data from Princeton and Stanford freshmen with data from other freshmen lends strength to this probability. The former group was separated from the latter because its members were known to have had somewhat more training in geography. In the case of two items, the percentages of correct performance in the higher and lower ability groups were the same. In forty-six cases, all other cases save two, the percentages of correct reactions by the group with more training were from 2 to 29 per cent higher than those by the other group.

Another procedure followed in seeking light upon the dependability of the exercises was the comparison of percentages of correct performance on various items by the *various* groups of college freshmen tested. A record of the reactions of the 1433 individuals comprising these groups was secured from the Commission. Except for the Princeton and Stanford groups, no data were furnished concerning the general abilities of these groups, the amount of geographic training they had had, and the character of that training. The amount and character of geographic training are known to vary greatly in different school systems. In some, there is no geographic training beyond that received in grade six; in others, none beyond grade seven; in still others, none beyond grade eight, and so on. The fact that these people had reached college level is the only indication furnished of their ability, and obviously, under the conditions noted, their geographic experiences doubtless had varied widely. No very close correlation in their products could be expected under these circumstances, but the accompanying graphs were constructed with a view to seeing what consistency, if any, there was in their performances. A general similarity is at once discernible. Although there was considerable range, as was to be expected, in the percentages on a given item, it is to be noted: that all groups performed best with the first third of the test, and least well with the last third of it; that high points in the various graphs occur in connection with the same items; and that in connection with another group of items low points occur in each. The general consistency of performance is striking and lends support to a belief in the dependability of the test.

PERCENTAGES OF CORRECT REACTIONS OF SIX GROUPS OF
COLLEGE FRESHMEN TO ITEMS IN PART II, FORM I
(Groups V and VI were advanced freshmen)





ANALYSIS OF RESULTS

A procedure commonly followed in determining the reliability of a given test is that of finding the coefficient of correlation between the scores of one hundred random samples of individuals on chance halves of the test. If the items in a test are all concerned with the same type of ability or with different phases of that ability which are of approximately the same difficulty and are functions one of another or of the same learning experiences, there should be marked positive correlation between these scores. If one thinks logically about such a procedure, however, he realizes that if a test deals with a wide range of abilities, some of which are known to be possessed by many, and others of which are known to be possessed by few—abilities which differ much as to the difficulty of developing them and as to the specific experiences necessary in so doing—little correlation should be expected between chance halves of the test as a whole. A low coefficient in this case would *not* mean unreliability. It is known, for example, that little correlation exists between one's ability to identify a strait or river or to read elevation from a color-contour map, and one's ability to read a polar projection, to read altitude from contours, or to translate distance in degrees of longitude into distance in miles at various latitudes. The only thing in common in the various abilities is that they relate to maps. They involve, however, the interpretation of entirely different symbols. It is known that nearly all people have had experiences which, however meager, lead to an ability to interpret some simple symbols. It also is known that the experiences of a vast majority of people have not been such as to lead to the ability to translate various other types of symbols. One logically could no more expect high correlation between performances on halves of a test which do not cover the same types of map-reading ability than he could expect it between two tests, one in American history and one in English history, if he knew that those tested had had experience with the one, and had not had experience with the other. Both tests are in history, but knowledge of American history does not breed knowledge of English history. Speaking mathematically, one is not a function of the other or of the same group of experiences. Two distinct

TESTS AND MEASUREMENTS

sets of experiences are needed to acquire knowledge in both fields of history. The same is true with reading semi-pictorial map symbols and other types of map symbols.

"In order to apply this rule of chance halves *in the only way which is sound*, then, the test had to be divided into groups of questions between which some degree of correlation might be thought, with reason, to exist. Then one could select *chance halves of each of these groups*. The halves of the test as a whole would thus consist of the chance halves from each group. Several exercises, for example, deal with getting the simpler types of ideas to be gained from a map network. Several others deal with getting the more complex types of ideas from a map network. Others deal with mere recognition of features from semi-pictorial symbols. Another group deals with reading various types of ideas from such symbols, and so on. From each of these groups, chance halves were taken. Even these halves were not such that an exceedingly high correlation logically could be expected. One may well read latitude without having the ability to relate latitude and crops, and these were grouped together. One may read latitude without being able to read longitude. One may be able to translate degrees of latitude into approximate distance in miles without being able to translate degrees of longitude similarly, and so on. The only halves of a test between which perfect correlation could be expected would be halves in which abilities were paired exactly.

There were not enough items in the test to permit such exact matching. Even with the rough matching possible, however, the coefficient of correlation between the scores of one hundred freshmen, selected at random from the 968, on the halves of the test consisting of chance halves from each group is .53. In view of the nature of the grouping, a correlation of .53 is satisfactory as evidence of reliability.

It is unfortunate that the plan of the committee for establishing reliability through the use of two or more carefully matched forms with the same groups was not effected. Since there was considerable range in the performance of different groups at freshman level

ANALYSIS OF RESULTS

with a given form, the performance of different groups with different forms gives an unsatisfactory basis for comparison.

It is significant, however, that the performance of different groups in given items in Form I, True-False, parallels rather closely that of different groups on other forms of those items. For example, the percentage of three college freshman groups reacting correctly to item 1, Form I, True-False, were 94, 93 and 95. The percentage of correct reaction on the part of all freshmen to this item was 94. This represents a group of 968. Form II, True-False, was tried on only 48 freshmen. Obviously, percentages computed on such different bases as 968 and 48 are of little value for comparison. The other forms, tried with 336 and 181 individuals afford better bases. The range of performance of the three groups *on the same item* (No. 1, Part II, T. F.) was 93-95. The range of performance of three other groups *on that item and corresponding ones in different forms* was 94-99. In the case of this particular exercise, the multiple choice form gives one somewhat more guidance in answering than does the true-false form. This fact, and the differences in the numbers of cases involved, doubtless account for the somewhat wider range on the different forms. In many other cases, the performance of three groups on the same question coincides closely with that of three other groups with three forms of the same question, as the following table shows.

Item in Form I True-False	Range of performance of three groups on the same item	Range of performance of three other groups on different forms of a given item
2	95-96	96-97
3	94-95	91-95
5	94-94	93-95
10	70-87	68-86
11	94-97	95-97
13	93-98	94-98
24	60-68	63-65
25	81-85	81-82
27	62-77	65-77
36	44-72	49-65
44	44-72	39-73
45	28-37	27-41

TESTS AND MEASUREMENTS

In some cases, the range of performance on the *same item* by different groups was *wider* than the range of performance by different groups on different forms of that item. For example:

17	84-93	90-95
21	76-85	78-79
23	69-80	70-73
28	41-57	45-51
30	79-89	80-83
32	64-72	59-65
37	43-56	51-62

The consistency of performance of six different, large groups indicated by these tables certainly is not a matter of chance. The fact that reactions to different forms show about the same results as reactions of different groups to the same form points strongly to the probability that, if the various forms had been given to the *same group*, there would have been very high correlation between the reactions of any given group on the several forms. If this be the case, then obviously the exercises are dependable and the differences in the performance of groups could be identified unquestionably with differences in their geographic abilities. Even in cases where the range was greater on the different forms than that on reactions of different groups to the same form, there is a marked coincidence of performance. For example:

6	61-70	57-74
8	85-92	77-93
9	51-73	58-83
14	82-87	74-85
15	71-77	58-75
19	73-77	69-75
20	64-66	50-65
34	62-67	50-65
38	11-14	6-22
39	54-71	40-65
41	44-57	37-56
48	47-75	35-75

ANALYSIS OF RESULTS

The findings from the foregoing considerations concerning the reliability of the test items in Part II may be summarized as follows:

1. The data furnished are inadequate as a basis for determining the true coefficient of reliability.
2. They serve, however, when handled *scientifically*, to establish the fact that the items have much reliability, since .53 is, under the conditions noted, a satisfactory coefficient of correlation.
3. Further evidences that the test has much reliability are:
 - (1) Close correlation between performance on it and *definitely known abilities* of two selected college groups tested.
 - (2) Close correlation between performance on it and abilities or lack of abilities known from investigations on the part of others.
 - (3) Performance by a "higher ability" college freshman group that proved consistently better than the performance of a "lower ability" group.
 - (4) A marked tendency toward consistency of performance on the part of six college groups, both on the same item and on different forms of that item.

There is no conclusive evidence of unreliability save in the cases of two items and possibly a third, to be noted later.

Statistics are of value to science only when used scientifically. When two factors are involved, as in the case of reliability, it is unscientific to base a conclusion on data dealing with only one of those factors. A valid conclusion can be reached only through careful consideration of both. In using chance halves of a test, selected without recognition of the known fact that abilities in the two halves are as different as abilities in two different fields, and in basing conclusions concerning reliability of the test on correlation between performance on those two halves, one is using only one factor and ignoring the other. One can not question the fact that the coefficient of correlation between *such* chance halves is low. Scientific thinking leads one to challenge, however, the statement in the report of the use of the test that this coefficient is a

TESTS AND MEASUREMENTS

coefficient of *reliability*. It is only a coefficient of correlation between two tests (considering each chance half as a test) on two different and only slightly correlated fields. The coefficient .53 undoubtedly is nearer the true coefficient of reliability than the coefficient .055 submitted to the committee in the report mentioned. It seems highly probable that, if complete data were in hand, a coefficient of reliability even higher than .53 would be established conclusively.

In connection with Part I, the same procedure was followed with similar results. The accompanying graphs¹ indicate the performance of various college freshman groups with this part. The coefficient of correlation between scores of one hundred freshmen, selected at random, on halves of the test consisting of chance halves of correlatable groups of items is .67.

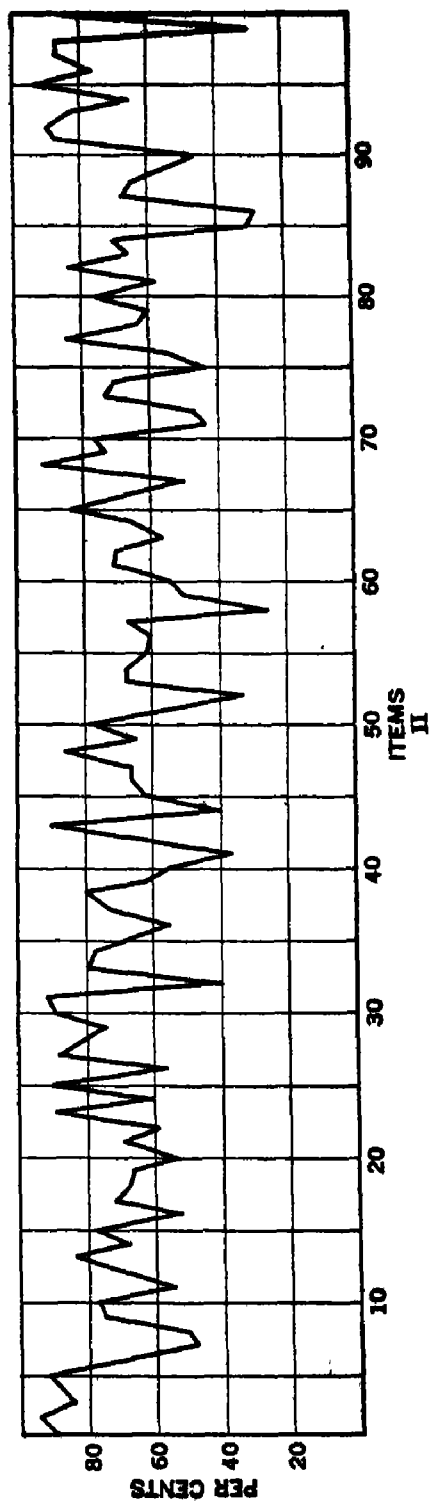
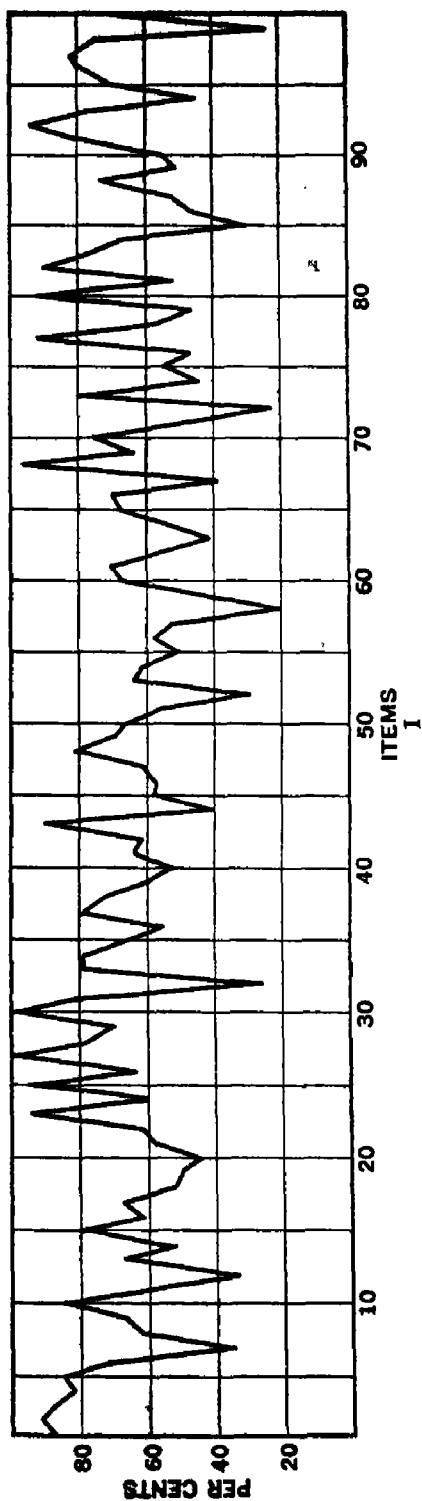
Mr. Floyd E. Masten, of the University of Chicago, who rendered valuable assistance to the committee in connection with the statistical work involved in the report, points out that this coefficient does not mean necessarily that Part I is more reliable than Part II, because it was based on a consideration of twice as many items. Since there are one hundred items in Part I and only fifty in Part II, halves of Part I were larger than those of Part II. Obviously, the coefficients .53 and .67 are not the true coefficients of reliability, for in neither case was the scope of abilities covered in the two halves of the test identical, but in all probability the true reliability coefficients, *which can be determined only after adequate data concerning both factors involved are in hand*, is higher rather than lower than these coefficients indicate. The foregoing considerations lead to the conclusion that the test is reasonably reliable as a *group* test for the purpose for which it was designed, namely, to determine what fundamental geographic abilities have been attained by students at various levels of school work.

Since in a given form an individual made only one reaction in the case of most of the abilities involved, it obviously is not possible with the use of a single form to tell which of the correct reactions on the part of that individual were a matter of knowledge

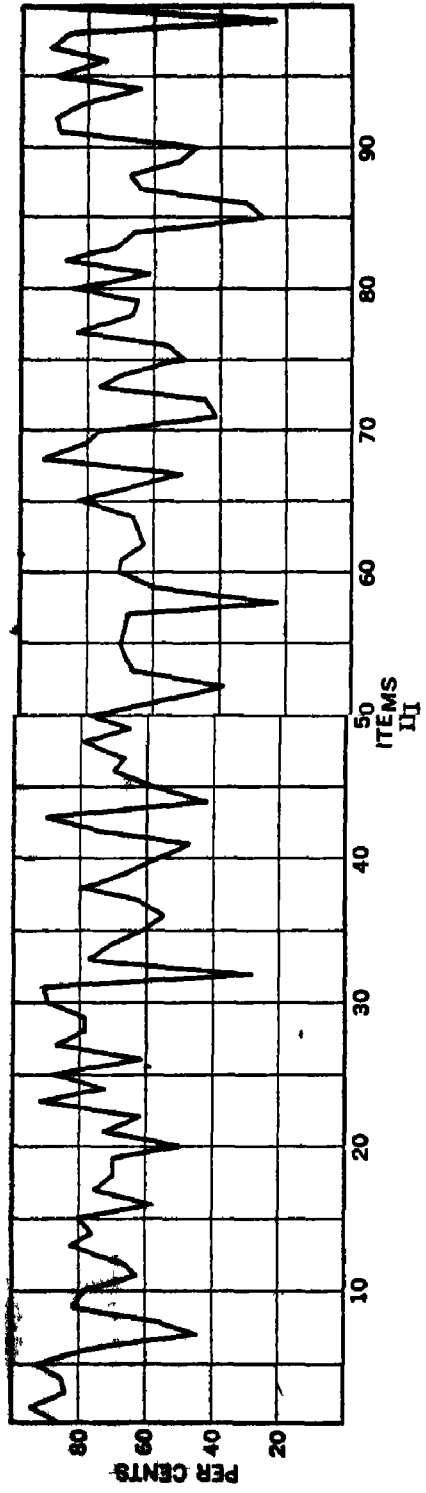
¹ P. 273 ff.

PERCENTAGES OF CORRECT REACTIONS OF SIX GROUPS OF COLLEGE FRESHMEN
TO ITEMS IN PART I.

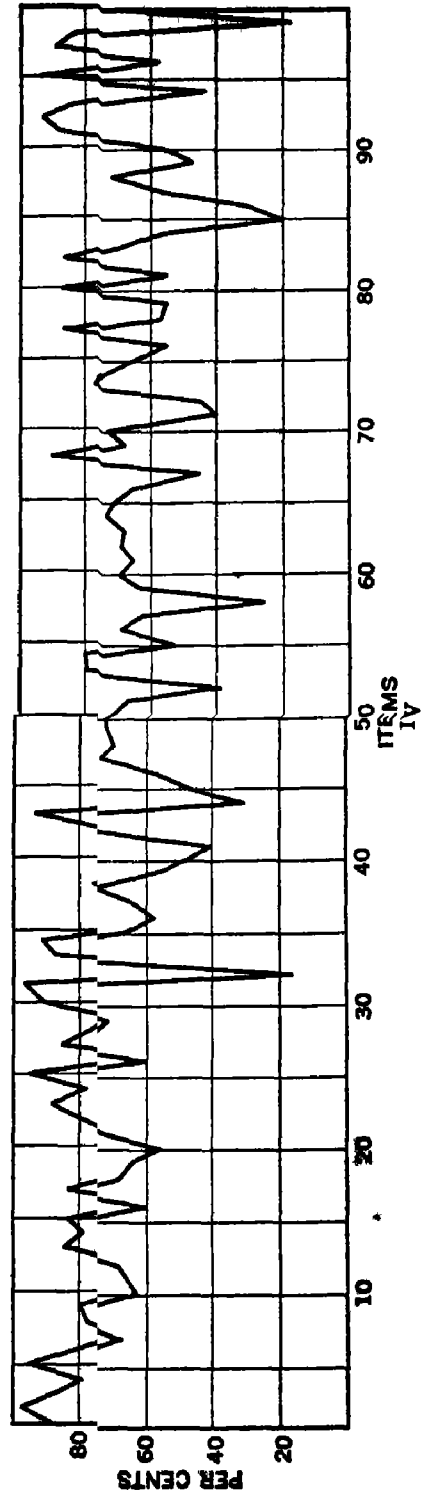
(Groups V and VI were advanced freshmen)



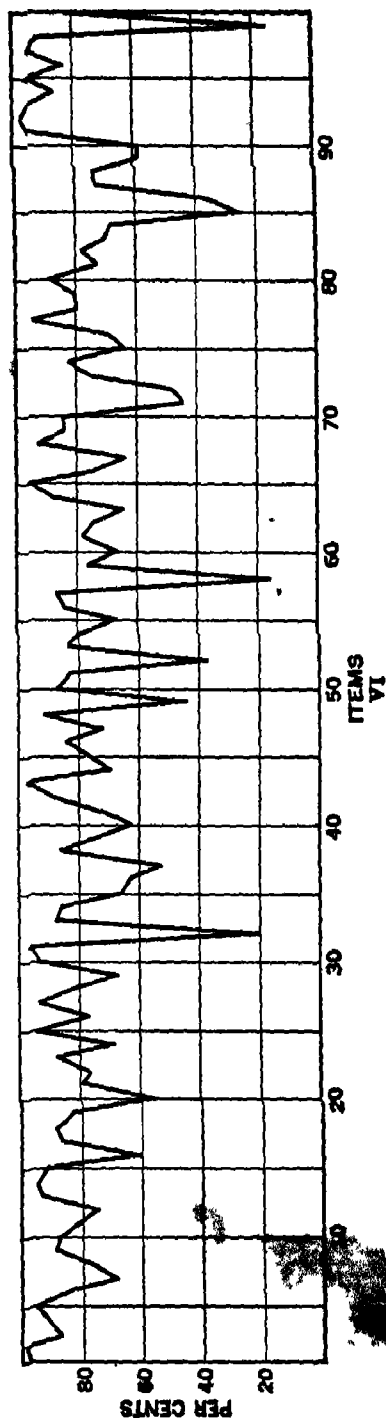
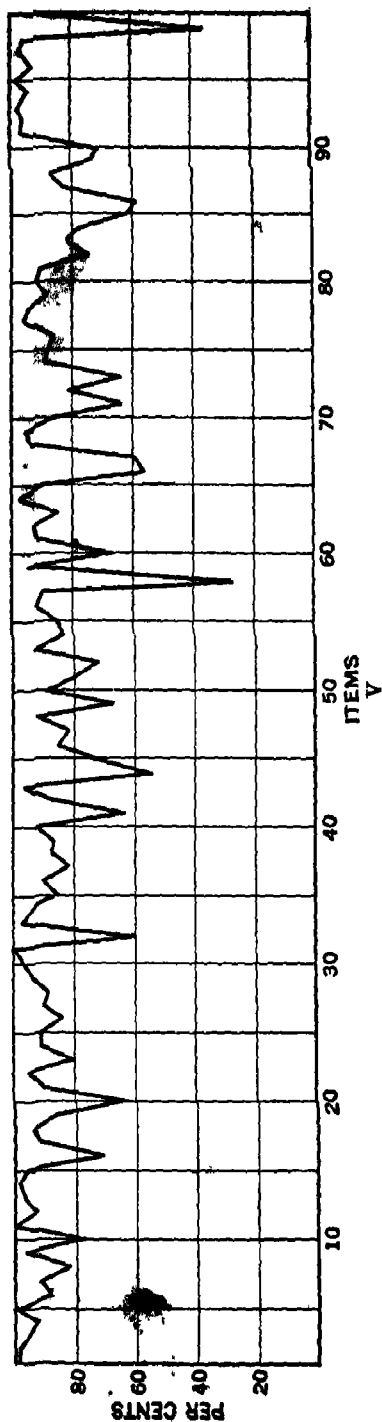
PERCENTAGES OF CORRECT REACTIONS OF SIX GROUPS OF COLLEGE FRESHMEN
 TO ITEMS IN PART I.—Continued
 (Groups V and VI were advanced freshmen)



[726]



PERCENTAGES OF CORRECT REACTIONS OF SIX GROUPS OF COLLEGE FRESHMEN
 TO ITEMS IN PART I.—Continued
 (Groups V and VI were advanced freshmen)



TESTS AND MEASUREMENTS

and which a matter of chance. If, however, all four forms are used with the same individuals, Part II is useful also in diagnosing individual cases.

Attention was turned next to the second question, namely, "In so far as the exercises are dependable, what do the results show about the attainments of the groups tested?" In order to have a convenient standard of comparison, it was decided to accept as satisfactory evidence of possession of a given ability on the part of a group the fact that 75 per cent or more of the group made correct reactions to the exercise which called for that ability. This choice, of course, was purely arbitrary. One might have chosen 70%, 80%, or some other at that general level. The only requisite was that the percentage chosen be high enough so it could not represent chance performance to any significant degree. The use of such a percentage would at least serve to separate abilities possessed by many from those possessed by comparatively few individuals.

The findings with regard to map-reading abilities were as follows:

Seventy-five per cent or more of the 225 sixth grade pupils tested reacted correctly to items 1, 2, 3, 4, 5, 11, 13 and 17 in Form I, Part II, True-False. No other form was tried with sixth grade classes. The abilities with which these items are concerned are the following:

Use of semi-pictorial symbols:

Ability to identify

- | | |
|--|-------|
| 1. land and sea..... | (87%) |
| 2. railroads..... | (90%) |
| 3. straits..... | (84%) |
| 4. cities and mountainous land..... | (85%) |
| 5. the southern versus the northern part of a given region..... | (89%) |
| 11. a gravel road versus other types of road (with aid of legend)..... | (85%) |

Use of color spot symbols:

- | | |
|---|-------|
| 13. Ability to identify land highest above sea level in a given region... | (85%) |
|---|-------|

Use of straight line network:

- | | |
|--------------------------------|-------|
| 17. Ability to read north..... | (77%) |
|--------------------------------|-------|

ANALYSIS OF RESULTS

It is to be noted that these eight abilities can be acquired more easily than any others involved in the test. It is in harmony with known facts about courses in fourth, fifth and sixth grade geography that pupils should have acquired them by the time they are in the sixth grade. As a matter of fact, the first six of them probably were acquired in the course of the fourth grade work and common intelligence makes it possible for one to tell with the aid of the legend which land in "Figure 4" is highest, even without previous experience with such a map.

Other abilities which were evidenced less conclusively, as the percentages in the parentheses indicate, were those involved in items 7, 8, 10, 12, 14, 15, 19, 21, 23, 25, 26 and 27, namely:

Use of semi-pictorial symbols:

Ability to get ideas of:

- | | | |
|-----|--|-------|
| 7. | relative location of river and railroad..... | (74%) |
| 8. | comparative distance without the use of a scale of miles..... | (71%) |
| 10. | relationship between gridiron street pattern and topography..... | (66%) |
| 12. | distance and highway marking as shown on auto road map..... | (61%) |

Use of color spot symbols:

- | | | |
|-----|---|-------|
| 14. | Ability to read and compare facts about density of population and amount of rainfall..... | (67%) |
| 15. | Ability to see probable relationship between facts read about crops and rainfall..... | (61%) |

Use of straight-line network:

- | | | |
|-----|--|-------|
| 21. | Ability to read direction from the equator..... | (65%) |
| 23. | Ability to use facts read about latitude in interpreting type of vegetation..... | (61%) |

Use of curved-line network:

- | | | |
|-----|---|-------|
| 19. | Ability to read direction..... | (62%) |
| 25. | Ability to read latitude in terms of degrees..... | (60%) |
| 26. | Ability to read latitude in terms of comparison with distance from equator to pole..... | (61%) |
| 27. | Ability to use facts read about latitude in interpreting length of growing season..... | (69%) |

The performance of the sixth grade with the test was almost exactly what was to be expected in the light of known conditions. It is to be noted that each of the second group of abilities involving the use of semi-pictorial symbols is of a higher order than any of those in the first group. In items 7, 8 and 10, comparisons were

TESTS AND MEASUREMENTS

involved, and in item 12 one has to deal with two types of figures on the map—those indicating mileage and those indicating highway numbering. Comparison also was involved in the exercises using color spot symbols, whereas in the one ability evidenced in the use of these symbols in the first group, no comparison was called for. It obviously is more difficult to read latitude and to relate it to vegetation and growing seasons than it is to read north on a straight line network. Nor is it surprising, in view of the maps in "Figures 8 and 9," that latitude was read correctly from the curved network (Item 25) in more cases than from the straight network (Item 22). Three facts make the latter a more difficult exercise. J in "Figure 9" is exactly on a parallel at the end of which is the number 60. In contrast, E in "Figure 8" is not on a parallel, and again the matter of comparison enters in. Secondly, all places in "Figure 9" are in north latitude, whereas all places in "Figure 8" are in south latitude except E. Thirdly, the convergence of meridians at the north pole in "Figure 9" is a clear indication of "northness" with regard to the equator. It at once suggests the north pole. In contrast, the only basis furnished in "Figure 8" is the numbering of the parallels and one must identify 0° with the equator and make comparisons based on that identification in order to realize that E is in north latitude. The point to be made is that the sixth grade pupils tested could read latitude, given the simplest means of so doing, but could not read it if these aids were not at hand. They had acquired only ability to use the more obvious signs of latitude.

In the case of the 29 items to which less than 60% of the sixth grade reacted correctly, the probability is that the abilities called for are not possessed generally by the group.

In the case of one item there is clear evidence that a 62% correct reaction does not evidence ability. This fact points to a weakness in the test. Item 42 was designed to call for the ability to read and compare range in temperature between two sets of places. Knowledge of experiences in elementary geography leads one to expect only chance correct answers to this item on the part of sixth grade pupils. This was the case with the other three exer-

ANALYSIS OF RESULTS

cises dealing with the ability to interpret isotherms, a fact which proved that this percentage meant either unusually fortunate guessing or some weakness in the exercise. Examination of the map in "Figure 13" on which the question was based shows that the number 54 is near G, the number 56 near H, and the number 64 near E. Though the figure near H does not represent temperature at H, the result gained from reading these figures instead of reading the isotherms gives the same result as would correct interpretation of the isotherms. Apparently, the children used these numbers instead of the isotherms. This possibility could have been obviated by omitting the numbers on the map on all isotherms except 64, 62 and 60.

Significant facts about the reactions of college freshmen (excluding the Stanford and Princeton groups) are the following:

1. Eighty-five per cent or more of the college freshmen reacted correctly to all eight items in the first sixth-grade list.

2. Eighty per cent or more of them also reacted correctly to items 8, 10, 14, 25, and 26 in the second sixth-grade list; seventy-five per cent or more of them to items 7, 15, 19, and 21 in that list; and, in addition, to items 30, 35, 48, and 49.

3. Whereas only eight items were handled correctly by three-fourths or more of the sixth-grade pupils, then, twenty-one items were handled correctly by three-fourths or more of the college freshmen. These twenty-one, moreover, *included the eight*.

4. Between sixty and seventy-five per cent of the freshmen reacted correctly to items 6, 12, 16, 20, 22, 23, 24, 27, 29, 32, 34, 36, 39, and 44.

5. Whereas twenty items, then, were reacted to correctly by sixty per cent or more of the sixth-grade pupils tested, thirty-five items were treated correctly by sixty per cent or more of the college freshmen, and *these included all the twenty items*. The difference between the forty per cent of the test handled fairly satisfactorily by the sixth-grade pupils and the seventy per cent of it handled equally well or better by college freshmen indicates *clearly* that gain was made in map-reading ability between the sixth and thirteenth school years.

TESTS AND MEASUREMENTS

6. Greater average ability on the part of the freshmen as compared with that on the part of the sixth-grade pupils also is evidenced by the fact that freshmen percentages were in no cases as low as percentages on corresponding items in the sixth grade; and that in the cases of fourteen of the twenty items in the sixth-grade lists, percentages were from nine to twenty-two per cent higher.

7. Items to which *less* than sixty per cent of the college freshmen reacted correctly were those numbered 9, 18, 28, 31, 33, 37, 38, 40, 41, 43, 45, 46, 47, and 50. (Item 42 was not included in these comparisons because of the objection noted to it in the earlier discussion.) Item 18 deals with east-west direction on a curved line projection; item 28, with the translation of distance in degrees of latitude to distance in miles; item 31, with longitude in degrees as compared with distance between the prime meridian and the 180th meridian; item 33, with the comparison of areas of regions mapped on different projections; item 38, with the comparison of areas of regions in different latitudes mapped on a straight-line network; item 37, with the relation of the continent pattern to latitude and longitude; item 40, with recognition of meridians on a polar network; items 41-43, with reading isotherms; items 45-47, with reading isobars; and item 50, with the identification of topographic profiles. These obviously are the more difficult phases of map-reading ability represented in the test. If any gaps are left in the ability to read maps in common use as a result of elementary and secondary training, those evidenced by incorrect reactions to the items just listed are the ones confidently to be expected. This expectation is based on (1) the relative difficulty of acquiring the abilities in question and (2) established facts concerning current teaching procedures.

8. With the exception of item 9, then, performance on the test coincides with the prediction of results on the part of the committee—a prediction based on definite knowledge of the two factors noted at the end of the preceding paragraph.

9. Item 9 calls for the use of a scale of miles. Evidence of lack of the ability involved was surprising. It may mean that, though the tool is introduced at early levels, practice in its use is insufficient

ANALYSIS OF RESULTS

to establish the habit of using it accurately. Perhaps many of those tested merely measured by eye comparison and did not trouble to measure the distance in the proper way. This possible explanation is in accord with the fact that performance was better on the corresponding item in the other forms. There is the third possibility that the term "about" was interpreted more liberally by some than by others. If so, the weakness is in the statement itself. The matter cannot be settled without further experimentation.

10. Summarizing, the results of the use of the map-reading tests give a clear picture of the probable gaps in map-reading ability which are left at the end of secondary school training, as it is commonly administered, and give a very definite notion of the type of advancement made in reading maps in grades six through twelve. The tests were, therefore, highly satisfactory for this one of the purposes for which they were designed.

The discussion of the value of Part I in answering the second major question, page 261, can be handled more briefly in connection with the discussion of the third question, namely: "What do the results [of the use of the tests] reveal concerning progress at various levels toward the goals involved?" Facts about known conditions that must be considered carefully as constituting one type of factor requisite to sound conclusions in the matter are the following:

1. There is no uniformity in the experience provided in so-called geography courses beyond the sixth grade, and little uniformity in those provided earlier.

2. Some of the abilities involved are those which tend to be developed, as one progresses toward maturity, through chance experiences, or through experiences in other fields. Any intelligent twelfth grade child who has had an effective course in European history, for example, would react correctly to more items in these tests than would a child of equal intelligence who has had other experiences like those of the first child but has had no such history course.

TESTS AND MEASUREMENTS

3. Reading habits of individuals vary greatly, and the more widely an individual reads, the more he tends to glean information which contributes to geographic abilities.

4. Many abilities tend to atrophy with disuse. Assumption that any given ability is possessed by a greater number of people in each successively higher grade ignores the significant factor of the degree to which the ability has been used in the interim.

5. Groups in the same grade vary greatly in general intelligence.

It also is to be noted that:

1. The percentages in the item profiles submitted are computed on bases ranging from 3 to 968. To use in an item profile percentages based on 3, 13, 26, 74, and 92 cases, in comparison with those based on several hundred cases, tends to give erroneous impressions.

2. If what a given group does in the sixth-grade is compared with what *that* group does in the seventh grade, the eighth grade, and higher grades, or with what is done at higher levels by another group *known* to have had the same learning experiences in grades below the sixth and to possess the *same* intelligence, the factors of difference in experience and intelligence would not need to be considered in interpreting a profile. The profiles submitted to the committee, however, *are not of this type*. They compare what one group did in the sixth grade with what a *different* group did in the seventh grade, a *third* group in the eighth, and so on.

By reason of the facts listed above about variations in groups, a variable which is exceedingly erratic must be taken into consideration in any sound interpretation of these profiles. In the report of the use of the tests to the committee, the profiles are interpreted as if they were profiles of the first type described; the second variable involved is ignored.

3. If profiles of the second type noted "fail to show appreciable grade-to-grade improvement" and "a large number of mean scores of items" are "of the same order of magnitude as chance scores," the items in question are classed in the aforementioned report as "poorer" items. It is also stated that such profiles indicate that the items "in question will not function as a measuring device for

ANALYSIS OF RESULTS

the grades in question." As well might it be said that a yard stick is not a good device for measuring the height of children (a variable which *tends* to increase from year to year) if one happens to use it on a group of children who, for some reason such as under-nourishment or thyroid abnormality, had not made the growth to be expected under good conditions.

Such reasoning concerning item profiles assumes that promotion from one grade to another inevitably is accompanied by appreciable improvement in the *abilities in question*, whatever the intelligence of the groups involved and whatever the nature of their learning experiences. This assumption is unquestionably false, and reasoning based on it therefore must result in unsound conclusions.

Any item is a good "yard stick" for measuring progress if reactions to it are consistently in accord with *known* abilities on the part of those who react. In the report, the *unquestionably* reliable item, "The distance from J to M is about 1400 miles" is marked "poor" as a measuring device because 47 per cent of the sixth-grade pupils tested reacted to it correctly, 52 per cent of the seventh-grade pupils, 45 per cent of the eighth grade pupils, and so on. What does nearness to the chance score mean? It means inability on the part of the majority of people in all three groups. One who *knows* does not resort to guessing. The yard stick shows, in other words, that children of all three groups are not as tall as they should be under conditions of proper nourishment. The school physician who made such a finding would not question his yardstick. He would try to find out what facts about the diet of these groups or about other functioning conditions of growth accounted for the retarded growth of individuals in the groups involved. Knowing beyond question that one who has the ability to read latitude and to translate degrees of latitude into approximate mileage reacts correctly with consistency to this item, one cannot doubt the *item*, and inevitably concludes that relatively few members of the groups tested had that ability.

Examination of current teaching practice shows that experiences of which this ability is a function are being provided erratically or not at all in elementary and secondary training. With these con-

ditions known, it is clear that, in so far as the sampling reported is adequate, the profile indicates *reliability* of this item as a measuring device, because its use results in the only type of outcome which possibly could be in accord with indisputable facts about functioning school experiences. One can safely maintain, accordingly, that it is not the *item* which is poor; instead, the *people tested* were poor in this ability. If this item is "poor" for the reasons cited, then a physician's thermometer is in order and reliable *only* when it shows that a patient has normal temperature; the thermometer is out of order whenever the temperature it registers is not what it would be if conditions that function in producing temperature were perfect.

Considering another aspect, assume that every fourth-grade child can identify a tree in a given picture. So can every fifth-grade child, every sixth-grade child, every seventh-grade child, and so on. Now according to the line of reasoning used in classifying the test items in question into "better" and "poorer" categories, this procedure is a poor device for finding out whether or not a child can recognize a tree in a picture, because there is no grade-to-grade improvement. It so happens that once one learns to identify in clear cut illustrations objects that are unmistakably trees, he has that ability to as great a degree as he could have it later. This ability doesn't improve from year to year. In the case of such ability, uniformity of attainment is the only result that possibly could be in harmony with the other factor involved in the correlation. If ability is constant, the correlated factor—test outcomes—must be constant, if the test item is reliable.

In physical tests, one does not conclude that a physician's thermometer is a poor measuring device because all healthy children have the same temperature. Taking temperatures is not ruled out as a valuable device for testing *one* aspect of one's physical condition by reason of this uniformity. It is considered important to know that the child's temperature is what it should be, if functioning conditions are right.

ANALYSIS OF RESULTS

4. Testing for map-reading ability, like testing for physical fitness, involves testing for several aspects of the matter. A child's medical chart may show his vision normal, his hearing defective, his height and weight satisfactory, his temperature normal, his chest expansion poor, and the like. Physical condition is rated in terms of the *number and nature of aspects in which his condition is satisfactory* as compared with the number and nature of the aspects in which it is unsatisfactory. In rating map-reading ability, the case is similar. One must, in measuring progress in it, deal with comparisons of the results of the test as a whole, as well as with profiles of individual items. The report to the committee takes no account of this second means of evidencing progress. Progress most certainly is measured by *a test as a whole*, if reactions to items 1, 2, and 3 (known to be reliable items) are made correctly by a high percentage of *sixth-grade* children and of those of *all higher grades* and if, further, few sixth-grade children react correctly to items 4, 5, 6, 7, and 8, while a high percentage of *seventh-grade children* and of those of *all higher grades* react correctly to them. This progress, moreover, would be evidenced conclusively *even if the profiles of all eight items were horizontal lines*. Sound conclusions concerning the value of a test as a device for measuring progress cannot be based on partial data. Progress must be viewed in terms of *new* abilities acquired at successively higher levels as well as in terms of an increase in the percentage of persons who have acquired given specific abilities.

In view of these considerations, the following illuminating tables were compiled from the item profiles submitted. In considering them, it should be kept in mind that the sampling of several groups was inadequate, and that wider use of the tests might alter the tables. However, the high degree to which the findings accord with predictions based on knowledge of current learning experiences leads one to believe that wider sampling would not alter them significantly.

In these tables is shown the performance of those tested in the sixth grade, the seventh grade, and the eighth grade, that of college

TESTS AND MEASUREMENTS

freshmen other than the Stanford and Princeton groups, and that of college freshmen in the latter groups. These are the only levels at which it is reasonable to expect progress to be evidenced *generally*, and these are the only groups for which the performance of more than one hundred and fifty pupils was reported. In the vast majority of cases, geography in the high school consists of one semester's work, which is elective for most students, which may be taken in any one of the three or four years of the high school course, and which deals largely in some cases with physiography, not geography, and in other cases chiefly with the production, consumption, and movement of commodities important in world trade—not with the geographic individualities of countries. Under these conditions, the only comparison with regard to Parts I and II which one could expect to evidence progress in high school would be comparison between eighth grade and freshman performance.

The only significant percentages in the profiles, then, are those for the five groups indicated. It must be remembered, too, that the college freshman group of the country as a whole is a more select group than the eighth grade group as a whole, because there is a tendency for the less proficient to be culled from the ranks by that time. Progress evidenced in terms of percentages of groups would tend, therefore, to be evidenced even if the freshmen had had no high school geography. Offsetting this, however, is the fact that it has been several years since most freshmen have had any geographic training, and that what they know evidences what they have retained in the interim, what they have strengthened by use in other fields, or what they have picked up by chance. The matter is by no means so simple as an item profile would make it seem, if one is interpreting the situation *scientifically*, and *not* on the ground of partial data.

Items in Part II, Form I, to which 75% or more of the sixth grade group and of groups at four higher levels reacted correctly

ANALYSIS OF RESULTS

Item	Percentages of Correct Reactions in Grades				
	6	7	8	13	13+
1.....	87	88	90	94	98
2.....	90	91	(91)	95	98
3.....	84	86	91	95	98
4.....	85	86	93	94	98
5.....	89	95	(94)	95	98
11.....	85	92	(92)	96	99
13.....	86	93	(93)	95	97
17.....	77	83	90	(90)	95

Additional items in Part II to which 75% or more of the *seventh* grade group and of groups at three higher levels reacted correctly.

Item	Percentages of Correct Reactions in Grades				
	6	7	8	13	13+
8.....	..	81	85	92	96
14.....	..	79	88	(85)	94
15.....	..	76	(75)	(75)	88

Additional items in Part II to which 75% or more of the *eighth* grade group and of groups at two higher levels reacted correctly.

Item	Percentages of Correct Reactions in Grades				
	6	7	8	13	13+
7.....	80	(78)	94
10.....	79	86	98
25.....	76	82	94
26.....	82	(82)	84
35.....	77	79	82

Additional items to which 75% or more of the college freshman and advanced freshman groups reacted correctly.

Item	Percentages of Correct Reactions in Grades				
	6	7	8	13	13+
19.....	75	94
21.....	78	92
30.....	83	91
48.....	75	84
49.....	80	85

TESTS AND MEASUREMENTS

Additional items to which 75% or more of the advanced freshman group reacted correctly.

Item	Percentages of Correct Reactions in Grades				
	6	7	8	13	13+
6.....	94
12.....	75
18.....	83
20.....	92
22.....	88
24.....	80
29.....	81
31.....	79
32.....	84
33.....	77
36.....	80
39.....	81
41.....	78
50.....	81

Items which failed to elicit correct responses from 75% or more of any group were 16, 27, 28, 34, 37, 38, 40, 43, 44, 45, 46, and 47. Item 23 was omitted from the eighth grade list, where it probably belongs, because though 77% of the eighth grade group and 94% of the advanced freshman group reacted correctly to it, the freshman percentage was not quite 75%—73%.

Tables of the foregoing types for Part I have been combined in the following table:

Items	Percentages of Correct Reactions in Grades				
	6	7	8	13	13+
1.....	83	88	(87)	90	98
5.....	89	90	93	(93)	97
13.....	75	(75)	80	83	94
23.....	75	85	(84)	90	(86)
25.....	82	86	(86)	91	95
30.....	80	81	82	90	93
34.....	80	(80)	(75)	(80)	87

ANALYSIS OF RESULTS

Items	Percentages of Correct Reactions in Grades				
	6	7	8	13	13+
2.....	..	80	84	95	99
9.....	..	83	(81)	(78)	90
17.....	..	76	79	88	(88)
28.....	..	76	(75)	79	82
38.....	..	76	79	(74)	86
48.....	..	83	85	(80)	92
80.....	..	78	85	(82)	89
14.....	76	(73)	95
27.....	81	89	93
31.....	91	92	97
43.....	83	91	97
82.....	80	86	(78)
92.....	78	91	97
97.....	80	88	96
98.....	79	86	93
100.....	77	82	86
3.....	85	90
4.....	85	91
6.....	75	85
10.....	75	86
15.....	81	92
33.....	81	89
50.....	77	87
65.....	81	94
68.....	92	93
70.....	76	85
73.....	75	(74)
77.....	85	95
91.....	84	95
93.....	80	96
95.....	92	96
8.....	77
11.....	85
12.....	79
18.....	88
19.....	84
21.....	81
22.....	80
26.....	78
39.....	87
42.....	90
45.....	76
46.....	84
51.....	82
53.....	87
54.....	80

TESTS AND MEASUREMENTS

Items	Percentages of Correct Reactions in Grades—(Continued)				
	6	7	8	13	13+
56.....	86
57.....	87
59.....	81
61.....	80
62.....	79
69.....	86
74.....	84
78.....	83
79.....	83
81.....	76
87.....	76
88.....	77
94.....	88
96.....	86

Items to which 75% or more of no group reacted correctly are 7, 16, 20, 24, 32, 35, 36, 41, 44, 47, 49, 52, 55, 58, 60, 63, 66, 67, 71, 72, 75, 76, 84, 85, 86, 89, 90, and 99. Twenty-eight per cent of the test, then, is approximately the gap left in the best current achievement represented. In the case of item 37, seventy-five per cent of the seventh grade reacted correctly and lower percentages of the higher groups did so. On item 40 the eighth grade performed better (76%) than the two higher groups. On item 64, the eighth grade performed better (75%) than the freshman group, but not so well as the advanced freshmen. On items 29 and 83, the freshmen performed better than the advanced freshmen. It is remarkable that such cases were not more numerous when one considers known differences in training of different groups and differences in the time between the date of training and that of taking the test. The fact that such knowledge and abilities as have been added at one level are retained or increased in most cases at higher levels certainly indicates progress.

The percentages in parentheses indicate the only items on which progress is not suggested by increased percentage of right reactions as well as by the additions evidenced.

Since Part I seems reliable, it would seem from this analysis that children are leaving the eighth grade with relatively little

ANALYSIS OF RESULTS

fundamental geographic attainment of the type for which that part tests. This is to be expected from analysis of current teaching procedures. Memorizing isolated facts does not make for ability to use them effectively and to *think* in terms of them. Whether the failures indicate lack of knowledge of facts or inability to use them in the simple new situations set up in the tests is a matter of little moment. What matters greatly is inability to do the latter. It is to be noted that:

1. Of the nine items concerned with recognizing signs of location with regard to the equator (Group I items) the eighth grade performance was good with only one—the simplest of them all.
2. Even the college freshmen did well with only two others of Group I—the next simplest two. The advanced freshmen did well with three others, and three in the group were not handled well by any group.
3. Of the forty-five items in Group II, the eighth grade handled well only one-fifth of them, identifying given conditions correctly with central Africa, the Congo, southern Europe, the Sahara, southern United States, eastern and western United States, New England, China, and Atlanta, Georgia.
4. College freshmen showed little more ability of this type, identifying given facts, in addition, only with southernmost North America, northernmost South America, and Lowell, Massachusetts. The advanced freshman group reacted well on two-thirds of the items in Group II.
5. With the other form of matching employed in items of Group III, the eighth grade handled well only six of the twenty-nine items involved—again about a fifth of the total. All these had to do with London or with colonizing countries of Europe. The college freshmen performed well with an additional eight, and the advanced freshmen with four more than the freshmen. The freshmen, then, handled approximately half the items of this group and the advanced freshmen about three-fifths.

TESTS AND MEASUREMENTS

6. Of the sixteen items in Group IV, evidencing knowledge of very simple relationships, a high percentage of eighth grade pupils reacted correctly to seven (2, 5, 9, 23, 27, 28, and 43); the college freshmen to those seven and three others (6, 29, and 77); and the advanced freshmen also to 8, 51, and 74.
7. The highest percentage of correct reactions to item 75, which calls for the mere reading of a rainfall graph, was 69, the advanced freshman percentage.

Summarizing, the analysis of the data furnished concerning Part I gives a clear, albeit distressing, picture of the paucity of geographic outcomes which are resulting from current procedures; it shows which types are resulting and which are not; and it provides information concerning the levels at which such knowledge and power as do result are being acquired.

In answering the second and third of the questions raised on page 261, the fourth question has been answered in essence. Efforts of remedial types should be directed toward filling the gaps in geographic achievement which the results of the tests disclose. Such light as the analysis sheds on the last question will be noted in the concluding section of the report.

CONCLUSIONS AND RECOMMENDATIONS

The committee realizes fully that the tests prepared by it represent only an initial step in the task of measuring the results of geographic training in elementary and secondary schools. This beginning, however, seems to have been effective for the purpose for which it was made. As to the tests themselves, four of the one hundred and fifty items in Form I should be revised. The difficulty with item 42 in Part II and the means of remedying it have been noted. Item 7 in Part II also should be changed because the map as it was printed may leave one in some doubt. Experimentation proved that the negative form of statement in items 41 and 42 of Part I gave some difficulty. These are the only negative statements in the test and they should be made positive. It was of value to discover, however, that even with advanced students, negative

ANALYSIS OF RESULTS

statements are confusing. Further experimentation is needed, as noted earlier, with the wording of item 9 in Part II. It is possible that in some cases the grouping of several items under one statement in Part I led those tested to resort to a multiple choice form of selection instead of reacting to each numbered item individually, as they were told to do in the initial directions. If this proves, upon further experimentation, to constitute a difficulty, it would be advisable either to make a separate statement for each, or to stress even more in the directions the fact that all the items under a given statement may be true, that all may be false, or that some may be true and others false.

After such revisions as seem appropriate in the light of the foregoing comments are made, further data should be collected from the use of the tests. The use should be such that adequate data for evaluating them as to reliability will result therefrom. While the data furnished to the committee do not show unreliability, they are not sufficient to do more than indicate considerable reliability. In order to handle statistically *and scientifically* both types of factors involved in establishing reliability, scores must be compared with intelligence test scores, reading scores, and the like, and with data concerning the specific amount and nature of the geographic training of those tested. In cases where there are not already strictly comparable items in two or more forms, such items should be supplied. It would be highly desirable to construct a battery of tests, one dealing with each phase of ability. A test of twenty items all dealing with the ability to read latitude from various types of projections, for example, could be so constructed and used as to establish beyond any doubt the value of items used for this purpose.

Judging from the results secured from the use of Parts I and II, it seems just as well that the expense of administering similar tests for Parts III and IV was not borne *at this juncture*. Since there is fully as much emphasis on mere fact memorization in courses currently administered in "Economic Geography" and "Principles of Geography" as there is in those in elementary geography, the findings undoubtedly would have been of the same general type.

TESTS AND MEASUREMENTS

Unsatisfactory achievement would have been revealed and it probably would have been practicable, as in the case of Parts I and II, only to have collected a small part of the data needed for their thoroughgoing evaluation. However, as soon as it is practicable to do with Parts III, IV and V what has been suggested with regard to the use of Parts I and II, they should be constructed and so used.

In constructing Parts III and IV, the committee planned to proceed much as with Part I. In Part III, instead of matching geographic characteristics with given regions, facts about the nature, conduct, and pattern of a given activity were to be identified with the activity by the use of matching devices similar to those used in Groups II and III, Part I. Items in Group IV, Part I, were to have had their counterpart in items which stated given aspects of an activity and called for the selection of human and natural conditions with which they are associated. The geographic characteristics of trade regions, ports, and trade routes, regions of production, and regions of consumption were to have been identified with specific trade, production, and consumption regions instead of with countries. In Part IV, corresponding groups of exercises were to have been provided for identifying various human and natural facts with given types of topography and topographic regions, different types of climate and climatic regions, and the like. No test gives a comprehensive view of fundamental geographic achievements if it does not test for all the major phases of such achievements.

It seems that the portions of the tests completed scarcely could fail to call additional attention (1) to types of outcomes which are of basic importance in geography, and (2) to the fact that many of these outcomes are *not* resulting from procedures now in common use. If this is the case, two of the chief purposes of the test will have been served.

It also seems clear that the testing devices used *have much reliability* and with further work along the lines indicated will constitute even more effective tools of measurement. To those who derive comfort from comparing accomplishments, not with

ANALYSIS OF RESULTS

what it is practicable and valuable to accomplish, but instead with what other folk have accomplished, the analysis of results affords a basis for the purpose. A given sixth grade performance can be compared with that of the sixth grade group tested, a given seventh grade performance with that of the seventh grade group tested, and so on.

So long, however, as one bases his procedure solely on what has been accomplished generally in the past his efforts toward improvement resemble trying to elevate one's self by tugging at one's bootstraps. If, instead, one sets achievements which are the *best* instead of the *most general* as his goal, his efforts resemble the use of a ladder in climbing. Though tests based on fundamental, practicable achievements show gaps in the outcomes of past training in geography which are not pleasant to contemplate, the results of their use should be a challenge, to all concerned, to direct efforts toward filling them. Only through effort directed toward achieving *the best* can many of the potential contributions of geography, to the effective pursuit of social studies become actual contributions. It is to the interest of those concerned primarily with the social studies, then, as well as of those concerned primarily with geography, to encourage and promote efforts directed to the improvement of geographic training.

SUMMARY OF REPORT UPON THE ITEM STUDY OF THE CALKINS-PARKER GEOGRAPHY TESTS

TRUMAN L. KELLEY

The full report by the writer upon the study of the items comprising the geography test, which is referred to by Miss Parker, contains item profiles for 300 geography test exercises. These call for far too much space to permit of their reproduction here, so that the essential portions of the report and a sample only of these item profiles are here given, following the next paragraph.

After a careful reading of the report by Miss Parker, the writer notes with regret that there is evidence of unfortunate misunderstanding as to the nature of evidence upon reliability and validity. Nevertheless it is not such as would lead him to modify his appraisals as to reliability and validity based upon the item study.

The basic data from which this report is made are the item profiles and table herewith of reliability coefficients. The reliabilities reported are those pertaining to single items of the test. They were obtained by splitting the items of a given class into halves, of, say, m items each, and correlating the halves for the groups as described. The result is a reliability coefficient for a test of m items. Then by means of the Spearman-Brown formula the reliability coefficient for one item has been estimated and is the figure given in the table upon the following page.

The item profiles show much absence of improvement from grade to grade in the abilities measured by a majority of the items, and they also show a large number of mean scores of items of the same order of magnitude as chance scores. Three sample profiles are here shown, (1) the "best" item, (2) an item of median excellence, and (3) the "poorest" item, of Form I, Part I.

Where either of these two facts is established for an item, it indicates that the item will not function as a measuring device for the grades in question. The items which fail to show appreciable grade-to-grade improvement are those listed as "poorer" items. Reference to the table shows that, with the exception of Form I, Part II, Stanford and Princeton Freshmen (where the

CALKINS-PARKER GEOGRAPHY TESTS

RELIABILITY OF ITEMS OF CALKINS-PARKER TEST AT DIFFERENT GRADE LEVELS

Part	Grade 7 (1)		Grade 9 (2)		Grade 13 (3) College Freshmen		Grade 13 (4) Stanford and Princeton		Graduate Students (5)	
	Better Items	Poorer Items	Better Items	Poorer Items	Better Items	Poorer Items	Better Items	Poorer Items	Better Items	Poorer Items
Fm. I, Pt. I, T. F.....	(a) .041	(b) .008	(a) .025	(b) .005	(a) .033	(b) .027	(a) .039	(b) .002
Fm. I, Pt. II, T. F.....	(c) .025	(d) .014	(c) .027	(d) .008	(c) .055	(d) .015	(c) .034	(d) .052
Fm. II, Pt. II, T. F.....	(e) .003	(f) .021	(e) .000	(f) .070
Fm. I, Pt. II, M. C.....	(g) .057	(h) .023	(g) .053	(h) .028	(g) .148	(h) .120
Fm. II, Pt. II, M. C.....	(i) .047	(j) .016	(i) .079	(j) .087	(i) .289	(j) .131

- (1) Random sample of 100 7th grade pupils, except for Form II, Part II, T. F., where a class of 25 was used.
- (2) Group of 98 for Parts I and II, Form I, of 41 for Form II, Part II, T. F., and random sample of 100 for other parts.
- (3) Random sample of 100 college freshmen.
- (4) This group (of 100 taken at random) is separated from other freshmen because the members had pursued specific courses covering certain phases of geography.
 - (a) Group of 70 for Form I, Part II, M. C. Group of 27 for Form II, Part II, M. C.
 - Items 1-4, 7, 24, 35, 36, 38, 47, 48, 50, 57, 69, 70, 73, 74, 77-79, 84, 87, 91, 96-98, 100 vs. 10-12, 15, 17, 22, 27, 29-31, 42, 43, 51, 61, 62, 64, 65, 68, 80, 82, 92-95.
 - (b) Items 5, 6, 8, 9, 23, 25-26, 33, 34, 37, 39, 40, 49, 56, 58, 71, 72, 75, 76, 85, 86, 89, 90, 88 vs. 13, 14, 16, 18-21, 28, 32, 41, 44-46, 52-55, 59, 60, 63, 66, 67, 81, 83.
 - (c) Items 9, 10, 11, 18-23, 29, 30, 32, 39, 40, 48, 49 vs. 1-6, 8, 14, 17, 25, 26, 35-37, 44, 45.
 - (d) Items 12, 13, 24, 31, 33, 34, 41, 42, 50 vs. 7, 15, 16, 27, 28, 38, 43, 46, 47.
 - (e) Items 1, 5, 6, 7, 29, 32, 39, 45, 50 vs. 12, 14, 16, 26, 28, 34, 37, 43, 44, 48.
 - (f) Items 2, 3, 4, 8, 17, 21, 22, 23, 24, 30, 40, 42, 47 vs. 9-11, 15, 16, 19, 20, 33, 35, 38, 49, 31.
 - (g) Items 7, 8, 9, 10, 15, 16, 24, 29, 30, 36, 38, 39, 47 vs. 5, 6, 11, 12-14, 21, 28, 32, 34, 40, 43, 49.
 - (h) Items 2, 1, 17, 18, 23, 25, 31, 37, 46, 48 vs. 1, 4, 19, 20, 22, 26, 33, 41, 42, 45, 50.
 - (i) Items 6, 7, 13, 14, 21, 22, 27, 28, 31, 34, 37, 38, 44, 50 vs. 4, 10, 11, 17, 18, 24, 25, 29, 30, 35, 36, 41, 42, 48.
 - (j) Items 1, 2, 8, 12, 15, 20, 32, 33, 43, 45, 49 vs. 3, 5, 9, 16, 19, 23, 26, 39, 40, 46, 47.

TESTS AND MEASUREMENTS
ITEM PROFILE
CALKINS-PARKER GEOGRAPHY TEST
FORM I, PART I

Grade	6	7	8	9	10	11	12	13	14	15	16	*
No. of Cases	225	301	193	131	74	92	26	968	13	3	74	465
99												
96												
93												
90												
87												
84												
81												
78												
75												
72												73
69								69				
66												
63									62			
60												
57							56	58				
54												
51												
48												
45												
42												
39												
36												
33												
30												
27												
24												
21												
18												

Characterized by T. L. Kelley as an excellent item.

*Stanford and Princeton freshmen.

item reliability is .052); Form II, Part II, True-False, Grade 9 (where the item reliability is .070); Form I, Part II, Multiple Choice, Graduate Students (where the item reliability is .120); and Form II, Part II, Multiple Choice, Grade 13 (where the item reliability is .087), and Graduate Students (where the item reliability is .131), these items fail to measure anything consistently for the grade in question. It is necessary to conclude that these items are not functioning as measuring devices for the elementary school, for the high school, and for the university, except in the exceptional cases noted.

CALKINS-PARKER GEOGRAPHY TESTS

ITEM PROFILE—*Continued*

Grade No. of Cases	6 225	7 301	8 193	9 131	10 74	11 92	12 26	13 968	14 13	15 3	16 74	* 465
99												
96												
93												
90												89
87												
84					84							
81						77						
78		75									76	
75												
72												
69			67	66				68	69			
66												
63												
60							58					
57												
54												
51												
48												
45												
42	44											
39												
36												
33												
30												
27												
24												
21												
18	Characterized by T. L. Kelley as a mediocre item.											

*Stanford and Princeton freshmen.

Though the "better" items show grade-to-grade improvement, this is not marked, and the item reliabilities remain so low in the 7th grade and in the 9th grade (except for Part II, Multiple Choice) that even these items can hardly be said to be serviceable in measuring individual or class differences at the elementary and junior high school levels. The "better" items, and as noted some of those listed as "poorer," do function to establish group and even individual differences at the university level, though Part I items are not functioning very efficiently.

It may be concluded that the test, when the better items only are retained, could be used to establish the inability of pupils as now found to cope with such problems, but it could not be used

TESTS AND MEASUREMENTS

ITEM PROFILE—*Continued*

Grade	6	7	8	9	10	11	12	13	14	15	16	*
No. of Cases	225	301	193	131	74	92	26	968	13	3	74	465
99									100			
96												
93												
90												
87												
84						83						
81											80	
78				78								
75	73	75	74		73		73					
72												
69								69				
66										67		
63												
60												59
57												
54												
51												
48												
45												
42												
39												
36												
33												

Characterized by T. L. Kelley as an item having no diagnostic power.

*Stanford and Princeton freshmen.

to establish individual differences as between pupil and pupil, nor even in any very serviceable way as between class and class at the junior high school level. It may well be that the items showing the shortcoming mentioned do measure a function which it is important to teach, and which, having been taught, could then be measured by means of these items. No evidence at hand confirms or refutes this hypothesis. The evidence indicates that as geography instruction now stands and as sampled by our testing program, the items do not reveal differences in ability.

Part II, Multiple Choice: The "better" items of the multiple choice test yield low but serviceable reliabilities for purposes of group measurement. In the 9th grade the item reliabilities are .052 (the average of .057 and .047). To yield a serviceable group test, the reliability should be .40 or better. Fifteen of the "better" multiple choice items will give a test of reliability .45, and as 53

CALKINS-PARKER GEOGRAPHY TESTS

such are available, it is possible to build up three forms of 15, 16, 17, or 18 items each, which would be serviceable for group measurement at this grade level. Probably the 18 item test would also be serviceable at the 7th grade level, as, judging by Part I, the reliability at this level will probably not be greatly different from that found at the 9th grade level.

Part II, True-False: The items of this part show no promise of yielding a group test for grades 7 and 9 under present conditions of geography instruction. It is probable that a serviceable group measure at the university level exists in these items, but as Form II, Part II, T. F., was not given at the university level no differentiation of the better and poorer items for the true-false part of this form is possible.

Form I, Part I, True-False: The average item reliability of the "better" items is .034 (average of .0141, .025, and .039). We therefore require a test of 23 items to yield a reliability of .45. As there are some 83 of these "better" items in Form I (Form II, Part I has not been tried out), it is entirely feasible to build up three forms of 25 items each which would be serviceable for group measurement. These forms would be serviceable for such measurement in the junior high school, the senior high school, and at the university level. The three together would yield a reliability of .71 which is scarcely reliable enough for purposes of individual diagnosis.

This report is complete for the data at hand, but due to certain gaps in the data, particularly as it concerns parts not tried out, and in view of the fundamental issue as to the value and place of a test which may measure important functions not now taught, but which does not well measure present attainment of pupils, no recommendation is made as to the next steps to be followed.

THE CONSTRUCTION OF EXERCISES IN THE USE OF HISTORICAL EVIDENCE

By MARION CLARK

Director of Upper Elementary and Junior High Schools,
Montclair, N. J.

This statement will cover six phases of the work done in the construction of two groups of exercises in the use of historical evidence, an elementary group and an advanced group: (1) the definition of the purpose and range of the tests, (2) the selection of materials suitable for these exercises, (3) construction and forms of the exercises, (4) developing directions for testing, (5) the testing of 2000 pupils and students, (6) statistical results.

The Exercises in the Use of Historical Evidence were made under the direction of Professors Truman Kelley, Henry Johnson, and A. C. Krey, during the years 1928-1930.

1. The specific purpose of the exercises is to test the ability to make legitimate use of a given piece of historical evidence, and of that only. The scope is very wide, since it includes historical material from the beginning of time to the present. The range is also wide, since the exercises are designed to test the ability to use evidence, from the third grade level to the graduate college level.

2. The first six months were spent in the collection of material selected from historical sources, from the works of historians, from maps and pictures. Certain difficulties and limitations were met.

First, all questions must necessarily lend themselves to a system of checks as answers, in order to make possible the scoring of the papers on an objective basis by clerical assistants.

Secondly, as far as possible, varying degrees of acceptability in answers must be avoided. This called for an absolutism of response, which is difficult to arrive at in the field of the social studies.

Thirdly, the vocabulary must be suited to successive grade levels.

The selection of material for the third grade, the lowest level to be tested, offered exceptional difficulty on account of the limited

HISTORICAL EVIDENCE

reading ability of the children. Even the Thorndike Word List was inadequate as a guide, for the reasons that it shows the *frequency* of words rather than their ease of reading and that it does not include on a third grade level many words such as *Eskimo* and *colonial*, which become familiar to children through social studies activities, while *airplanes*, of course, does not appear in it at all. Also, it is difficult to make exercises on this level which call for some simple type of historical thinking beyond the mere power to read and interpret.

The selection for the highest levels offered the problem of arranging adult materials on levels of increasing difficulty.

By October of 1929 a sufficient number of exercises had been worked out to call for approximately 2,000 responses. In the first form all of the exercises called for conclusions to be drawn from evidence. Critical analysis and suggestions for reorganization of the material were then undertaken. Professor Henry Johnson, and others in the field of history, suggested the rejection or revision of some material improperly used as evidence. Dr. Truman Kelley criticized the exercises from the standpoint of their suitability as testing material; and a number of laymen who read the exercises critically offered helpful suggestions.

3. As a result of conferences with Professors T. L. Kelley, Henry Johnson, and A. C. Krey, the exercises were recast into two distinct tests; the first for children on an elementary school level from the third through the sixth grade, the second for students in the junior high school, high school, and college.

The advanced test was divided into four groups: Group I was designed to test the ability to draw conclusions from evidence; Group II was designed to test the ability to discover one fact from another known fact; Group III was designed to test the ability to group evidence that is pertinent to a problem; Group IV was designed to test the ability to discriminate among types of evidence.

An attempt was made to organize the elementary test into the same four groupings. However, no terms equivalent in meaning to "general fact," "specific fact," and other types of evidence were

TESTS AND MEASUREMENTS

found within the vocabulary possibilities of third and fourth grade children. Therefore, no Group IV was developed in the elementary exercises.

To provide for historical thinking for each response, it was desirable to lessen the possibilities of guessing. This was secured by two means: (a) the use of three, instead of two, possible choices for each response; (b) the use of an unequal number of statements wherever two lists were to be matched.

The tests in this form were criticized again by Professor Henry Johnson and reorganized.

Before the exercises were published it was necessary to develop directions for pupils to be printed at the beginning of each exercise. Vocabulary difficulties were numerous. Several months were spent in trying out various wordings. The elementary tests were tried by twenty individual children who commented upon their understanding of the directions. The advanced test was given to several adults of outstanding critical keenness, in order that inconsistencies and irrelevancies might be discovered. Among these adults were Miss Dorothea Marston, Director of the Child Guidance Department of Montclair, New Jersey; Miss Margaret Mathias, author of *Beginnings of Art in the Public Schools*, *Art in the Elementary Schools* and *The Teaching of Art*; Miss Pearl Monks, principal of Wade Park School, Cleveland, Ohio; and Dr. Mary Reed of Teachers College.

A final conference with Dr. A. C. Krey and Dr. Kelley on October 4, 1930, resulted in further revisions, and the decision that the exercises should be printed for experimentation. The tests¹ at that time consisted of:

ELEMENTARY EXERCISES

Group	Number of Exercises	Number of Responses Called for
I. Drawing Conclusions from Evidence..	8	37
II. Making Proper Implications from Facts.....	5	18
III. Grouping under General Heads.....	5	34

¹The tests from which the tables accompanying this paper have been drawn are published by Charles Scribner's Sons.

HISTORICAL EVIDENCE

ADVANCED EXERCISES

Group	Number of Exercises	Number of Responses Called for
I. Drawing Conclusions from Evidence..	9	55
II. Making Proper Implications from Facts.....	8	22
III. Grouping under General Heads.....	7	47
IV. Classifying Types of Evidence.....	6	30

Examples of these exercises with a study of statistical results are included at a later point in this article.

4. The next step in the procedure was to develop directions for giving the tests. Two thousand booklets of each group of exercises had been printed. The exercises were given to four high school classes, four junior high school classes, and six elementary classes by testers who adapted their explanations to the needs of each class as the tests were given. A stenographer took careful notes of all that was said by the tester, questions that were asked by the children, and observations in regard to various types of difficulties that the children had with the questions, exercises that the children especially enjoyed and any other observations that might be of help in organizing the final directions for giving the test. From these notes, directions for giving the tests were drawn up, submitted to Dr. Kelley, and revised upon the basis of his criticisms.

5. The remaining tests (approximately 1,500 of each exercise) were given with the assistance of the following groups: Professor Henry Johnson of Teachers College, Columbia University; Mr. Frank G. Pickell and the principals and teachers of the Montclair Public Schools; Professor Roy Hatch and Miss Florence Stryker of the State Teachers College at Montclair; Miss Anna Gannett, principal, and the teachers of Brookside School in Montclair; Miss Clara Levy, Miss Eda Willard and Miss Rachel Jarrold, heads of the departments of history of the state normal schools of New Jersey at Newark, Glassboro, and Trenton, respectively; Mr. Harold Ferguson, principal of the high school in Montclair and Miss North, the head of the department in history in that high school;

TESTS AND MEASUREMENTS

Mr. William Davis, supervising principal of Haddon Heights, New Jersey; Mr. Charles Beek, principal of the junior high school in Summit, New Jersey; Miss Jane Gilbert of the Buckley Schools in New York; Sister Mary Boniface, School of the Holy Child, Summit, New Jersey; Mr. Guy Baskerville, University School, Pittsburgh, Pennsylvania.

6. The papers were sent to Dr. Truman Kelley for examination and statistical analysis.

The number of pupils used in the various grade levels are:

Grade 3—176 pupils	Grade 11— 14 pupils
Grade 4—261 pupils	Grade 12— 22 pupils
Grade 5—266 pupils	Grade 13—482 pupils
Grade 6—297 pupils	Grade 14— 57 pupils
Grade 7—108 pupils	Grade 15— 35 pupils
Grade 8—107 pupils	Grade 16— 21 pupils
Grade 9—110 pupils	Grade 17— 21 pupils
Grade 10—23 pupils	

For convenience, the detailed discussion of the exercises and statistical analysis is divided into two parts: 1. The discussion of the elementary exercises. 2. The discussion of the advanced exercises.

ELEMENTARY EXERCISES

For purposes of illustration, exercises have been selected containing items that showed promising results, and items so lacking in promise that they were discarded.

GROUP I

EXERCISE IV

This is the only book the little children in America had in early colonial schools. It was called a horn book. On a thin piece of wood, usually about four or five inches long and two inches wide, was placed a sheet of paper a trifle smaller than the wood. At the top the alphabet was printed in large and small letters. Below were simple syllables such as ab, eb, ib, etc. Then came the Lord's

ELEMENTARY EXERCISES

Prayer. This printed page was covered with a thin piece of yellowish horn which showed the letters through so they could be read.

After you have looked at the picture and read the story about it carefully, place a check in the proper square after each statement.



	True	Does not Tell	False
1. The children in colonial schools did not learn to read.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. The people in colonial days thought it was important for children to learn their letters.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The paper used in the horn books was made in America.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Many interesting stories were written for the boys and girls in the American colonies to read in school.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. There were printing presses in those days.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

TESTS AND MEASUREMENTS

The profiles of the items in this exercise follow:

ELEMENTARY—GROUP I EXERCISE IV

	Item 1 Grades				Item 2 Grades				Item 3 Grades				Item 4 Grades				Item 5 Grades			
	3	4	5	6	3	4	5	6	3	4	5	6	3	4	5	6	3	4	5	6
99
96
93
90
87
84
81
78	79	79
75	74
72
69	68
66
63
60	..	61
57	56	55
54
51	52
48	48
45	43	43
42	41
39	40
36
33	32
30	30
27	27
24
21	19
18	17
15
12	11	..
9	8
Mean per cent correct	255/4				223/4				195/4				145/4				65/4			
Sum of mean per cents correct=883/20=44.15																				

In the case of items 1, 2, 3 and 4, the improvement from grade to grade is evident and regular. Item 5, however, shows a lower number of correct responses, on each successively higher grade level. This is somewhat puzzling, since one would expect children to assume that printed pages used in school, proved the presence of printing presses of some sort. The suggestion is that thoughtful

ELEMENTARY EXERCISES

children would decide that there was nothing to correspond with our present printing press. It is, perhaps, even more likely that the children looked for a definite statement that there was such a press, and, not finding it, marked the answer, "Does not tell." The item could not be included in the revised form of the test.

GROUP II

The behavior of the items in all the exercises in this group was so good, that I have chosen to present here an overlapping exercise, that is, one which was used in both the elementary and the advanced tests.

EXERCISE IV

These facts all concern the making of the first railroad across the United States. It was built by two companies, the Union Pacific and the Central Pacific.

Fact I. Congress gave about \$32,000 and a large piece of land for each mile of railroad which either company built.

Fact II. One day in 1869, 4,000 laborers, mostly Chinamen, laid more than ten miles of track.

Fact III. Sometimes Indians surprised the builders and drove them away.

Fact IV. Not long after the railroad was finished herds of buffalo were no longer seen crossing the tracks, but great herds of cattle grazed on the plains instead.

Put a check in the proper square under the number of each given fact which helps you to know the first statement. Do the same with each of the other statements.

	Fact I	Fact II	Fact III	Fact IV
1. Congress was not trying to help one company to beat the other..	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. In 1869 Chinese laborers were not forbidden to come into the United States as they are now.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The United States government thought it was very important for the east and the west to be joined by a railroad.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. The railroads were built under great difficulties.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Some of the people who moved west undertook cattle raising...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The behavior of Item 3 is given here in detail to indicate the general type of result obtained in this exercise; namely, an excellent profile through the ninth grade, but no discriminating value on the higher levels.

TESTS AND MEASUREMENTS
ELEMENTARY—GROUP II, EXERCISE IV, AND
ADVANCED—GROUP II, EXERCISE I

ITEM 3

	Grades																
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		
99.....	
96.....	
93.....	94	
90.....	
87.....	86	86	87	89	..	86	..	
84.....	84	
81.....	
78.....	78	76	
75.....	
72.....	71	
69.....	
66.....	
63.....	
60.....	
57.....	58	
54.....	
51.....	
48.....	
45.....	47	
42.....	≡	
39.....	
36.....	
33.....	
30.....	30	
27.....	≡	
24.....	
21.....	..	23	
18.....	..	≡	
15.....	
12.....	
	11	
	≡	

Mean per cent correct = .986/15

Some of the items in this exercise showed a good profile only through the sixth grade. No items had differentiating value for high school or college students.

ELEMENTARY EXERCISES

Dr. Kelley's reports on Group II of the Elementary Exercises was as follows: "As all of the exercises look promising, the reliability of the entire test for 100 fourth grade pupils was calculated. The test was split in two parts, Exercise II and IV being one part, and I, III, and V the other. The correlation between these parts is .527 so that the reliability of the entire test is .69. To secure a group test of reliability .40 or better, two exercises having a total of 6 items or more are required. We therefore need but 4 of the 5 exercises in order to secure two forms of the desired reliability. I suggest that Form I be made up of Exercises II and V, and Form II of Exercises III and IV. The reliability of either of these in the fourth grade is approximately .48, which is very satisfactory for a group test."

Group III

All but one of the 34 items in the five exercises in this group showed good profiles, although Exercises I and II were too easy, as shown by a profile range approximately from 72 to 97 per cent correct.

Exercises IV and V were used as overlapping tests, being the last two of the elementary exercises and the first two of the advanced exercises.

Exercise IV is given here:

If the Spanish king had tried to prove in 1620 that he owned all of North America, which of these facts could he have used? Place a check in the proper square at the right of each statement to show whether he could or could not.

	He could have used these facts	He could not have used these facts
1. In 1493 the Pope declared that all the land west of a certain line belonged to Spain. This gift included all North America.....	<input type="checkbox"/>	<input type="checkbox"/>
2. Columbus touched the mainland of America in 1598.....	<input type="checkbox"/>	<input type="checkbox"/>
3. The English king in 1606 gave the land along the eastern coast of North America to two companies, the London Company and the Plymouth Company.	<input type="checkbox"/>	<input type="checkbox"/>

(Continued on following page)

TESTS AND MEASUREMENTS

(Continued from previous page)

	He could have used these facts	He could not have used these facts
4. Columbus discovered America and claimed it all for Spain in 1492.....	<input type="checkbox"/>	<input type="checkbox"/>
5. In 1609 Henry Hudson discovered the Hudson River and claimed it for Holland.....	<input type="checkbox"/>	<input type="checkbox"/>
6. The English king said that England owned the land because John Cabot had discovered it about 100 years before.....	<input type="checkbox"/>	<input type="checkbox"/>
7. In 1524 Verrazano sailed along the coast of North America, entered New York harbor, and claimed it for France.....	<input type="checkbox"/>	<input type="checkbox"/>

ITEM I

	Grades															
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
99.....	
96.....	
93.....	93	95	..	95	95	
90.....	92	..	89	
87.....	≡	
84.....	
81.....	
78.....	77	78	
75.....	
72.....	70	..	70	
69.....	≡	69	
66.....	
63.....	64	
60.....	
57.....	
54.....	..	54	
51.....	
	(chance score)															
48.....	
45.....	
42.....	
39.....	
36.....	
33.....	34	

Mean per cent correct = 1170/15

ELEMENTARY EXERCISES

The profile of Item I is an illustration of an item that showed differentiating power from the third through the twelfth grade.

A summary of the findings on the elementary tests, and the recommendations for the making of the revised tests into Form A and Form B, are also given.

ELEMENTARY EXERCISES (1,000 pupils—Grades 3 to 6 inc.)

Group and Exercise	No. of Items in each Exercise	Mean per cents correct	No. of Items showing fair or good profiles	Items recommended for Revised Form A	Items recommended for Revised Form B	Reliability of two forms found by correlating scores from 100 4th grade pupils
Group I						
Ex. I.....	5	60.35	5	All items		
Ex. II.....	5	65.05	5		All items	
Ex. III....	3	48.00	2			.500
Ex. IV....	5	44.15	4	1, 2, 3, 4		(°A=3.34; °B=3.95)
Ex. V.....	5	55.05	2			
Ex. VI....	5	49.75	3			
Ex. VII....	5	62.63	5		All items	
Ex. VIII..	4	48.73	1			
Group II						
Ex. I.....	3	79.17	3			
Ex. II.....	3	66.25	3	All items		
Ex. III....	3	73.42	3		All items	.48
Ex. IV....	5	81.09	5		All items	
Ex. V....	4	78.13	4	All items		
Group III						
Ex. I.....	8	86.96	8			
Ex. II.....	7	87.78	7	All items		.431
Ex. III....	7	77.14	7		All items	(°A=3.25; °B=3.05)
Ex. IV....	7	76.34	6	All items		
Ex. V....	5	73.60	5		All items	

The exercises, reset into Forms A and B according to these recommendations, have been published by Charles Scribner's Sons and are available through the publishers or through the American

TESTS AND MEASUREMENTS

Historical Association Investigation of the Social Studies in the Schools.

ADVANCED TEST

Group I (Drawing Conclusions from Evidence)

Two complete exercises are given here; Exercise IV because its results illustrate various unpromising situations; and Exercise VIII because it illustrates hopeful possibilities.

EXERCISE IV

During the Revolutionary War which was fought to defend the independence of America, the American army was often without proper supplies. In 1777, on the 22d of December, George Washington, the commander of the American army, was prevented from attacking Howe because two brigades of soldiers had refused to fight until they had food. They had been without bread and meat for two or three days.

Neither did the soldiers have proper clothing. When they marched into winter quarters in 1777, their feet were bloody.

During the same year the farmers in Pennsylvania had plenty of grain and vegetables to sell. At the very time when the American soldiers were suffering from winter cold in Valley Forge, there were hundreds of pairs of shoes, stockings and clothing lying in the woods for want of teams to draw them to the army and money to pay teamsters.

Congress, a body of men chosen by the different states, had the following plan for providing for the army. They appointed two commissary-generals, one to take charge of purchasing supplies and one to take charge of distributing them. However, the men who did the buying and transporting were responsible, not to the commissary-generals, but directly to committees in Congress. These committees discussed the matter and decided what money should be allowed and gave orders for carrying on of the work.

After you have read this paragraph carefully, place a check in the appropriate square after each statement.

	True	Not proved by evidence	False
1. That Congress made no effort to supply army with food and clothes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. That the organization of Congress into committees did not promote efficiency in the army.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. That there were no supplies to be obtained in the colonies.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. That Congress was entirely responsible for the discouraging conditions in Washington's army.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The profiles of Items 1 and 2 follow:

TESTS AND MEASUREMENTS

The profile of Item 1 shows clearly that the question is too easy for these levels. In a new try-out of these questions Item 1 might be stated: That Congress was more interested in politics than in supplying the army with food and clothes.

In this case the correct response would be, "Does not tell." It seems probable that this might not be so easy.

The profile of Item 2 shows sufficient difficulty and some discrimination of grade, but the results are so erratic that they suggest a weakness in the item. It would seem wise before trying the exercise again, to submit this item to the criticism of several history experts and to reword the item on the basis of their judgments.

The profiles of Item 3 and Item 4 are somewhat different from those of 1 and 2.

Item 3 has differentiating value on all three levels, junior high, high, and college, and could be retained in resetting the test. The peculiar behavior of the item on the eleventh and seventeenth grade levels requires a detailed re-examination of the papers and a study of causes.

Item 4 shows fair differentiating power up to the college level.

It is evident that Exercise IV should be reshaped and that further experimenting should take place. The nature of the changes which seem to be called for are discussed in a later paragraph.

EXERCISE VIII

In the year 1800 Spain ceded to France, New Orleans and the mouth of the Mississippi. Thomas Jefferson became president the following spring. He believed that this would interfere with the trade of the United States through the Mississippi, River. He appointed Mr. Livingston and Mr. Monroe to try to buy New Orleans from France. Napoleon, then dictator of France, offered to sell United States the land which Jefferson wanted and also the great tract of land called Louisiana. The powers of the government are defined by the constitution.

Three clauses in the constitution were interpreted by some people as covering the power of the United States to buy Louisiana.

Article 1, Section 8.—"The Congress shall have power to lay and collect Taxes, Duties, Imposts, and Excises, to pay the Debts and provide for the Common Defense and general Welfare of the United States."

* * *

"To make all Laws which shall be necessary and proper for carrying into Execution the foregoing Powers, and all other Powers vested by this constitution in the Government of the United States, or in any Department or Offices Thereof."

ADVANCED—GROUP I EXERCISE IV

[317]

TESTS AND MEASUREMENTS

Article IV, Section 3.—“The congress shall have Power to dispose of and make all needful Rules and Regulations respecting the Territory or other Property belonging to the United States.”

President Jefferson did not believe that these clauses gave the government the right to add new land. He said he thought that it was “Safer not to permit the enlargement of the Union but by amendment to the Constitution.”

When the treaty by which all of the land mentioned above came to Congress to be ratified, Jefferson wrote, “I suppose they (Congress) must then appeal to the nation for an additional article to the Constitution approving and confirming an act which the nation had not previously authorized.”

Jefferson prepared an amendment to the Constitution permitting the purchase of land. He thought this should be adopted. Some of the Senators did not think that was necessary but Jefferson said, “Our peculiar security is in the possession of a written Constitution. Let use not make it a blank paper by construction.”

The day after he wrote the above quotation, Jefferson had a letter from Mr. Livingston in Paris urging him to hurry the ratification of the treaty. Then Jefferson wrote another letter in which he said, “The less we say about constitutional difficulties respecting Louisiana the better.”

Within four days the treaty was ratified and the land was purchased, though no change had been made in the Constitution. When the president presented the treaty to Congress to be voted upon, he did not mention the constitutional difficulty. Later on, Senator J. Q. Adams proposed an amendment to the Constitution to authorize the addition of land. It was not seconded and has not been heard of since.

After you have read these paragraphs carefully, place a check in the appropriate square after each statement.

All the items, except Item 4, show power and might be retained. However, this exercise, in common with most of the exercises in Group I, show clearly that three problems must be taken into consideration before resetting the exercises. The first problem is that of setting up a series of items based upon given evidence, each item of which will show improvement in response from grade to grade, over a range of ten grades; that is, from grade VII to grade XVII. Certain probabilities appear: that the evidence, and the given items, difficult enough to challenge college students, would be so far beyond the reading comprehension of seventh and eighth grade pupils that the element of guessing would enter. The reverse is also true; that material easy enough for junior high school pupils to comprehend and yet difficult enough to test the power of college students to draw conclusions, is hard to find. The attempt to formulate the more difficult items for these exercises has resulted in some that are obtuse rather than clear and difficult. Thus the complication of various interpretations by ad-

ADVANCED EXERCISES

vanced students is added to the possibility of guessing by the junior students.

In remaking the test, there are two possible courses to follow. One possibility would be to divide the exercises into two parts, limiting the scope of each, one for grades seven to eleven, inclusive, retaining much of the present material, and one for grades twelve to seventeen, inclusive. The other possibility would be to expect each exercise to show differentiating power from item to item, rather than within each item.

The second problem is that the difficulty of shutting out all evidence except that bit which is presented, apparently increases with increased education. Some of the responses to items indicate that such is the case, and discussions with high school and college

	True	Not proved by evidence	False
1. The framers of the Constitution thought that United States in 1783 was as large as its successful development would ever permit.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. They did not think about the possibilities suggested by large pieces of land belonging to foreign countries lying next to United States territory.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. In case territory was to be added to the United States an interpretation of the Constitution or an amendment to it would have to be made.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Jefferson was insincere either in his belief that the purchase was unconstitutional, or in his oath of office to support the Constitution.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. When Jefferson saw the dangers of European occupation of the Mississippi Valley, he subordinated his preconceived views about the strict construction of the constitution and joined with the majority of Congress to seize the opportunity to open the continent to the expansion of the United States.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Jefferson changed his mind about the constitutionality of the Louisiana Purchase.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The profiles of this exercise are as follows:			

ADVANCED—GROUP I EXERCISE VIII

441/11

ADVANCED EXERCISES

ADVANCED—GROUP I EXERCISE VIII

	ITEM 3																	ITEM 4																
	7	8	9	10	11	12	13	14	15	16	17	Grades		7	8	9	10	11	12	13	14	15	16	17										
99										
..										
81	81										
78										
75	73	..	75	74										
72										
69										
66	68	67												
63										
60	..	62										
57										
54	54										
51	52										
48	50										
45										
42	..	43										
39										
36	36										
33	(chance score)																			34	33			33				
30	30										
27	27										
24	24	24										
21	21										
18										
15										
12	13										
Mean per cent correct = 699/11																	= 306/11																	

TESTS AND MEASUREMENTS

ADVANCED—GROUP I EXERCISE VIII

	ITEM 5																	ITEM 6																
	Grades																																	
	7	8	9	10	11	12	13	14	15	16	17	7	8	9	10	11	12	13	14	15	16	17												
99												
..												
87												
84	86													
81	83													
78	81													
75													
72													
69													
66													
63	65	65													
60	..	62	62													
57	59													
54	..	56													
51													
48	50													
45													
42													
36	40													
33 (chance score)												33																						
30													
27	27													
24													
21													
18	18													
..													
9	8													

Mean per cent correct = 709/11

= 336/11

ADVANCED EXERCISES

classes have brought to light that many of the more thoughtful students responded on the basis of previous information about the situation pictured. The question arises then, "Is it possible that in the matter of drawing conclusions from evidence, the less one knows, the better?" It might be helpful before resetting this group of exercises to give the test to two groups of students of comparable intelligence, one group with history training and one group without it. Some such procedure should throw light upon the interpretation of the behavior of the more difficult items.

The third problem to be taken into consideration concerns the amount of material that can be presented as evidence and yet be well-considered in the use of a number of items within a reasonable length of time. An obvious weakness in the exercises of Group I is that the constructor of the tests attempted to approximate the kind of problem which would be experienced in a "life situation." That is, enough material was presented as evidence so that some selection would be necessary. Many high school students who were given the exercises complained that they could not do justice to Group I within the period. Teachers remarked, also, that the less careful students "finished" the exercises first. The suggestion of the constructor of the tests is that the evidence be reduced to the smallest amount necessary to the consideration of the items, and that a few items which require an impracticable amount of evidence be excluded.

For example, in Exercise VIII, a brief statement, such as, "The Constitution does not state that the Congress or the President shall have power to purchase land, though some statesmen have believed this power may be implied from some general statements in the Constitution," might replace the quotations from the Constitution. The last two sentences of the evidence and possibly one or two others could be omitted.

A considerable space has been devoted to a discussion of the difficulties of Group I of the Advanced Test since this was the only group which showed such unpromising returns that Dr. Kelley considered it necessary to renew the entire attempt to build the exercises.

TESTS AND MEASUREMENTS

A more detailed description of the other exercises seems impracticable here, but those exercises are available through the Commission. The following statement summarizes the general conditions of this group of exercises.

It has already been stated in this article that the two overlapping exercises in Group I, that is, the last two of the Elementary Test and the identical exercises used as the first two in the Advanced Test, proved to be too easy for college levels and, in some cases, too easy for high school levels.

Of the 47 responses called for by the seven remaining exercises in Advanced Group I, 22 showed *fair* or *good* promise, 5 showed either a weakness in the structure of the Item or incorrect scoring, and the remaining 20 had insufficient discriminating power.

ADVANCED TEST

Group II (Exercises in Making Proper Implications from Facts)

Seven of the eight Exercises in Group II brought responses which made the exercises promising. All, except Exercise VI and Item 1 in Exercise VII, were recommended by Professor Kelley to be retained.

Exercise V, with its profile, is here given as an example:

EXERCISE V

In 1861 a civil conflict broke out between the government of the United States and a group of its Southern States which had seceded and formed the Confederate States of America with Jefferson Davis as President.

The following statements are quotations from English political leaders during that struggle:

Given facts:

- I. The London Times of October 8, 1862, quotes Mr. Gladstone, who was Chancellor of the Exchequer, as saying, "Jefferson Davis had . . . made a nation. . . . We may anticipate with certainty the success of the Southern States."
- II. Mr. Mason, who was the commissioner from the Confederate States to England, wrote, "It is perfectly understood in the House of Commons that the war professedly waged to restore the Union is hopeless and the sympathies of four-fifths of its members are with the South."
- III. Mr. Belmont, an American who was an agent for Rothschilds in England, says that he talked with the English Prime Minister, who said to him, "We do not like slavery, but we want cotton, and we dislike very much your Morrill tariff."
- IV. In 1863, John Bright, the labor leader, at an immense meeting of trades-unions said: "Privilege has shuddered at what might happen to old Europe

EXERCISE V

888/11

TESTS AND MEASUREMENTS

ADVANCED—GROUP II EXERCISE V

ITEM 3											
	7	8	9	10	11	12	13	14	15	16	17
110
109
...
106
...
103
102
99
96
93
90
87
84
81
78
75
72
69
66
63
60
57
54
51

Item 3: Fair.

Exercise entire is promising.

ADVANCED EXERCISES

if this grand experiment (of democracy in America) should succeed. But you, the workmen, you struggling upwards toward the light with slow and painful steps—you have no cause to look with jealousy upon a country which, menaced by the great nations of the globe, is that one where labor has met with the highest honor."

Place a check in the appropriate square, or squares, after each of the following statements to show which given fact, or facts, show this statement to be true.

	Fact I	Fact II	Fact III	Fact IV
1. Some English leaders were interested in the Civil War for commercial reasons.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Some labor leaders believe that the favorable attitude of the English leaders toward the South was due in part to their jealousy of the success of the American republic.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. English leaders seemed to think that the great military disadvantage accruing to the South because of the fact that they were in rebellion against a government already established and recognized was more than offset by various advantages.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Exercise VI, which failed to show promise is also given here:

EXERCISE VI

Given facts:

1. From 1821 to 1830 many people from the United States moved to Texas and were given land by Mexico.
2. The Texans, who were largely from the United States, fought for and won their independence from Mexico in 1836.
3. Texas asked to join the United States.
4. Texas people said they must be allowed to keep their slaves.

Place a check in the appropriate square, or squares, after each of the following statements to show which given fact, or facts, show this statement to be true. Statements:

	Fact 1	Fact 2	Fact 3	Fact 4
1. The request of Texas for admission to the United States presented this problem: whether it would be wise to have additional members of Congress who favored the extension of slavery.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Texas was an independent nation during a part of its history.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Mexico endangered the integrity of its own territory in its eagerness to settle its open lands.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

TESTS AND MEASUREMENTS

ADVANCED—GROUP II

EXERCISE VI

	ITEM 1										
	7	8	9	10	11	12	13	14	15	16	17
123
...
119
...
116
...
114
...
111
109
108
107
105
...	100	100	100
99	98
96	94	95	95	..	95	..
93	..	91	≡	..	94
90
87
84
81	81

Item 1: No differentiating power in this item.

ADVANCED EXERCISES

ADVANCED—GROUP II EXERCISE VI

ITEM 2																	ITEM 3																
Grades																	Grades																
7	8	9	10	11	12	13	14	15	16	17	7	8	9	10	11	12	13	14	15	16	17												
99												
96												
93	93												
90												
87												
84	..	82	84	83	..	82	85	86	..	86	86	86												
81	..	82	83	83	81												
78	80	79	..	78												
75	77												
72	71	71												
69	68												
66												
63	64	64												
60												
57	57	..												
54												
51												
Mean per cent correct = 903/11											842/11																						
Item 2: Negligible differentiating power in this item.											Item 3: No differentiating power in this item.																						

TESTS AND MEASUREMENTS

ADVANCED TEST

Group III

Of the 35 items in this group, 22 showed *fair* or *good* promise, 2 items gave indication of being incorrectly scored. The other 11 items showed too little differentiating power.

Exercise V is given as an example here and the profiles of Items 1, 2, 5 and 7, each of which illustrates a different type of behavior.

EXERCISE V

At the close of the Revolutionary War by which the United States gained its independence, some of the people of the United States were better off financially than they had been at the beginning of the war, though the government was in a desperate condition financially.

By checks placed in the appropriate squares, show which of the following statements would indicate improved financial conditions of certain groups, which would indicate government financial conditions to be in a desperate condition, and which statements you would reject as not showing either of these conditions.

Item 1 shows differentiating power up to grade fifteen. Its failure to continue on levels sixteen and seventeen raises again the question of greater difficulty for more advanced students whose thinking includes many other facts which suggest other possibilities than the one considered desirable here.

Item 2 shows the reversal of expectation, which indicates improper scoring of answers.

Item 5 presents the confused picture of an item with no differentiating power, and Item 7, the steady rise of a promising item.

The two forms A and B have been set up, but it would certainly appear that further experimental work should be done to secure a better test of this type. Two suggestions are: (a) that the whole of Group III be submitted to 15 college professors for their criticisms of material, phrasing of items, and desirable responses, and (b) that the exercises be divided into two groups, one for grades seven to eleven, inclusive, and one for grades twelve to seventeen, inclusive, as was suggested for Group I.

ADVANCED EXERCISES

	Indicates that some were better off financially	Indicates that government was in bad financial condition	Does not show either condition
1. During the war the fishing industry and the shipping business were destroyed.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. The foreign armies brought with them the coin of their various countries.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Congress issued a great quantity of paper money which depreciated in the hands of the people until it became valueless.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. When a foreign army was quartered for a long season in any place, the farmers in the neighborhood received better prices, and usually in specie, for their products.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. In June, 1783, a band of 80 soldiers marched to Philadelphia and demanded their pay at the point of the bayonet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. In 1781 Congress proposed an impost duty of 5 per cent on some articles in order to raise money to pay the national debt, but not all of the states consented and the project fell through.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. During the Revolution the country gained about 300,000 inhabitants.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. While the British army was quartered in the cities of Boston, New York, Philadelphia, they had spent money lavishly on entertainment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Congress had no power to collect taxes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

TESTS AND MEASUREMENTS

ADVANCED—GROUP III EXERCISE V

	ITEM 1																	ITEM 2																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
	Grades																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
	7	8	9	10	11	12	13	14	15	16	17	7	8	9	10	11	12	13	14	15	16	17																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
99</

ADVANCED EXERCISES
ADVANCED—GROUP III
EXERCISE V

	ITEM 5																	ITEM 7																
	7	8	9	10	11	12	13	14	15	16	17	7	8	9	10	11	12	13	14	15	16	17												
99	100											
96	95	96	95											
93	93	91	..	93	..	94											
90	91	90												
87	86	..	85	86	86												
84	83	85												
81	..	80	81												
78	79	76	..	78												
75												
72												
69												
66												
63												
60												
57												
54	56												
51												
48												
45												
42												
39												
36												
33	(chance score)																																	
	Mean per cent correct=957/11																																	
	=851/11																																	

TESTS AND MEASUREMENTS

ADVANCED

Group IV (To discriminate among types of evidence)

Of the 30 responses called for by these exercises, 23 showed *fair* or *good* promise, 2 showed evidences of improper scoring, and 4 showed slight differentiating power. Since these latter four were not needed in the revised test, no study has been made of their performance.

Exercise II is given here with its profiles.

EXERCISE II

The following statements concern President Johnson who succeeded President Lincoln.

	General fact	Specific fact	Writer's personal opinion of situation	Writer's personal opinion of motive
1. Early in life he had been a poor tailor.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. His administration would have been more successful if he had possessed Lincoln's tact.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Many of his recommendations proved to be very valuable to the nation.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Johnson kept most of Lincoln's cabinet because he wished to follow Lincoln's policies.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. When Johnson was impeached and tried by the Senate the vote for conviction was 35 ayes to 19 noes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

The question arises whether the exercises in Group IV are intrinsically better than those in Group I, or whether the brevity of the material permitted a more thoughtful handling of each item, thereby bringing more consistent responses.

A summary of the entire Advanced Test follows:

ADVANCED EXERCISES
ADVANCED—GROUP IV
EXERCISE II

ITEM 1																	ITEM 2																
Grades																	Grades																
7	8	9	10	11	12	13	14	15	16	17	7	8	9	10	11	12	13	14	15	16	17												
99												
96												
93												
90	90	90												
87	89												
84	85	86												
81	81	83												
78	79												
75												
72	72	74												
69	70	72												
66	64	65												
63	64												
60	62	63												
57	58												
54	..	56	57												
51	55												
48	50												
45												
42	..	44												
39												
36												

Mean per cent correct=718/11

Item 1: Good

Mean per cent correct=819/11

Item 2: Very Good

TESTS AND MEASUREMENTS

ADVANCED—GROUP IV

EXERCISE II

ITEM 3																	ITEM 4																
Grades																																	
7	8	9	10	11	12	13	14	15	16	17	7	8	9	10	11	12	13	14	15	16	17												
99	100												
96	94												
93													
90													
87													
84	84												
81													
78													
75													
72													
69													
66													
63													
60													
57													
54													
51													
48													
45													
42													
42													
Mean per cent correct=681/11																	=819/11																
Item 3: Good																	Item 4: Good																

ADVANCED EXERCISES
ADVANCED—GROUP IV
EXERCISE II

	ITEM 5 Grades										
	7	8	9	10	11	12	13	14	15	16	17
99
96	95
93	92
90	≡	89	89	90	..
87	86
84
81
78
75
72
69
66	65	65	64
63
60
57	..	55
54
51
48
45
42
39	38
36

Mean per cent correct = 828/11

Sum of mean per cent correct = 70.27

Item 5: Good

Exercise entire is very promising

ADVANCED TEST

(1,000 pupils—Grades 7-17 inc.)

Group and Exercise	No. of Item	Mean per cents correct	No. of Items showing fair or good profiles	Items recommended for Revised Form A	Items recommended for Revised Form B	Reliability of two forms, by correlating scores from 100 students
Group I						
*Ex. I.....	In this group of Exercises only Exercise VI and Exercise VIII, and seven individual items in other exercises showed sufficient differentiating power to recommend them for use in a revised form. Two attempts on the part of Dr. Kelley to construct revised Forms A & B resulted in correlations so low that he recommended rebuilding this test. (See above discussion of difficulties).		
*Ex. II.....			
Ex. III.....	5	75.62	..			
Ex. IV.....	4	64.11	..			
Ex. V.....	9	46.02	..			
Ex. VI.....	4	65.34	..			
Ex. VII.....	9	64.69	..			
Ex. VIII..	6	48.30	..			
Ex. IX....	10	41.50	..			
Group II						
*Ex. I.....
*Ex. II.....
Ex. III.....	3	77.70	3	All items	100 college students—.330**
Ex. IV.....	4	not computed	4	All items	($\bar{O}_A=3.06$; $\bar{O}_B=4.00$)
Ex. V.....	3	not computed	3	All items	100 eighth grade pupils—.577
Ex. VI.....	3	not computed	0	($\bar{O}_A=1.82$; $\bar{O}_B=1.87$)
Ex. VII.....	4	not computed	3	Items 2, 3, 4	
Ex. VIII..	3	not computed	3	All items	

*Exercises I and II of Groups I, II and III are overlapping exercises, being included in the Elementary Test as the last two exercises, and in the Advanced Test as the first two exercises. The per cents correct are included in the table of the Elementary Test.

**Professor Kelley's report states, "The correlation for the freshman level is a trifle low but hardly so low, but that the test can be used there, particularly in view of its excellence at lower levels."

ADVANCED EXERCISES

ADVANCED TEST—Continued

Group and Exercise	No. of Item	Mean per cents correct	No. of Items showing fair or good profiles	Items recommended for Revised Form A	Items recommended for Revised Form B	Reliability of two forms by correlating scores from 100 students
Group III						
*Ex. I.....	100 College freshmen—
*Ex. II.....30**.—Note: Evidences
Ex. III....	6	74.27	5	Items 1–6 incl.	of incorrect scoring of
Ex. IV....	8	69.36	4	Items 1, 3, 4, 6, 7, 8	Item 2 of Ex. V and Item
Ex. V....	9	71.54	6	4 of Ex. VII led to the
Ex. VI....	7	65.61	3	Items 3 and 6	recommendation that
Ex. VII....	5	47.20	3	Items 1, 2, 3	these be rescored, since
						this would bring about a
						higher correlation.
Group IV						
Ex. I.....	5	55.58	4	Items 4 and 5
Ex. II....	5	70.27	4	Items 1, 3, 4, 5
Ex. III....	5	52.09	4
Ex. IV....	5	52.53	2	Items 1 and 3	100 college freshmen—
Ex. V....	5	57.58	5363***
Ex. VI....	5	58.69	4	Items 1, 2, 3, 5

**Professor Kelley's statement regarding this correlation is, "This is a trifle lower than is very much to be desired. I would therefore recommend that Item 2 of Exercise V and Item 4 of Exercise VII be rescored and a recomputation of the reliability made, incorporating these items. This might raise the correlations to .4 for college freshmen, and, as the reliability is undoubtedly higher in the eighth grade than at the college level, we would then have two exercises of sufficient reliability to be generally serviceable."

***Professor Kelley's statement is, "This is a trifle low, but the test should still be serviceable for most group measurement purposes."

CHAPTER V

INTERESTS AND ATTITUDES

The planning committee recognized among the expected outcomes of instruction in this field points of view, interests, and attitudes. Its statement was:

Points of View, Interests, and Attitudes:

Points of view, interests, and attitudes are usually accepted as outcomes of the teaching of these subjects. All propaganda rests upon that belief. Such attitudes are difficult enough to define, and still more difficult to test or measure. A promising beginning has been made and it seems worth while to support further effort. Among the points of view, attitudes, and interests, are the following:

A perspective on current affairs. The task of affording pupils a basis for viewing events about them in a reflective manner is shared by nearly all of the social studies. This value arises more specifically from the comparisons with, or contrast to, similar activities in the past or in distant lands. Tests for this value must be constructed out of present situations.

Historical-mindedness. Usually characterized as the habitual association of social occurrences with the time at which they took place. It means, more fully, a recognition of society as a constantly changing complex, whose details vary as the complex changes.

Locational-mindedness. The habitual association of social occurrence with the place in which they occurred.

Concern for the common good. Patriotism is a more common designation of this objective. It involves training the individual to recognize his own welfare in the welfare of the group, or to respect the common welfare as a desirable end in itself.

Tolerance, racial, religious, national, social. Related closely to the preceding, it deals more specifically with certain prejudices which seem inherent and lead to injustice. It is believed that acquaintance with and understanding of other races, religions, parties, and social groups will lead to a fairer judgment of them, with resultant benefit to the individual and society.

Leisure time interests. Acquaintance with the social world, past and present, is commonly regarded as capable of arousing in the pupil a passion or interest to find out as much as possible about some congenial topic, or groups of topics, whose pursuit will afford him satisfaction and pleasantly occupy his leisure. With increased leisure a common condition, this objective becomes increasingly important. Tests must be devised.

The adviser on tests felt a special interest in some of the more fundamental aspects of the problem of attitudes. With M. R. Trabue, he undertook to contrive a test of certain personality traits, which is here reported. He also engaged A. M. Jordan to make a supplementary study of the relation of pupils' cheating on examinations to certain circumstances. This study is also reported

INTERESTS AND ATTITUDES

here.¹ The progress of attitude testing was so far advanced that it seemed best to engage L. L. Thurstone to construct tests of attitudes on certain questions. The Commission selected a series of issues for which Thurstone constructed tests. His technique, which had worked so well at adult levels, presented some difficulties when applied at the earlier grades of school life. These studies are discussed elsewhere in this volume.²

¹See page 437 ff.

²See page 83 ff.

THE MEASUREMENT OF CERTAIN TRAITS OF CHARACTER

TRUMAN L. KELLEY and M. R. TRABUE

SECTION I

HISTORICAL STATEMENT

The study of character has been and still is being attempted by techniques which, because of their lack of precision, were discarded long ago in such other psychological fields as psycho-physics, mental measurement, vocational abilities, scholastic achievement, and the like. The development of objective measures of character traits is a fundamental necessity in the improvement of methods of character training and in the remedial work connected with penology, delinquency, and other manifestations of maladjustment. Serviceable tools for measuring character status objectively must be developed before it will be possible to demonstrate convincingly the superiority of any method of character training or correction.

The analysis of Spranger and the methods of measurement developed by Strong, Allport, Bernreuter, Thurstone, and others give promise of instruments of measurement and analysis sufficiently radical, novel, and reliable to suggest the opening of a new field of psychology. As this tends to be accomplished, it may be expected that one attendant source of error in character measurement will grow larger and larger,—that connected with disingenuous responses. The temptation to give responses which indicate a socially approved character rather than those which indicate the true character of the subject will become increasingly powerful as more and more practical dependence is placed upon the results of these measurements.

Coaching is serious enough in connection with the measurement of achievement. Though a child could not take a Thorndike psychological examination and earn a college entrance score without actually knowing a good deal, by a study of the Thurstone

MEASUREMENT OF CHARACTER TRAITS

scoring key he might learn the "right" responses in an attitude scale, thereby making a higher mark, even though his character had been negligibly improved, or possibly changed in the opposite direction. The vocation interest scores resulting from taking the Strong Interest Blank are less likely to be distorted by falsification of the record because (a) a falsification characteristically will lower the interest score in one vocation while increasing it in another, which being of questionable advantage to the one taking the test constitutes a very weak stimulus to prevarication, and because (b) it is frequently difficult or impossible for the taker of the test to know whether a given falsification will have a desired result upon his score.

For quite similar reasons the authors believe that insincere responses will not seriously affect character indexes obtained from a free association test, wherein a stimulus word is presented and the subject responds with the first word that comes to mind. It is true that a subject, knowing that an "honesty" score is to be calculated, might intentionally alter all of his responses with intent to get a high score. However, since the association test as here developed and scored by the authors results in eight character measurements, rather than in one, and since a modification of responses to raise one character score would typically tend to alter, frequently to lower, other character scores, and since the subject is scarcely wise enough to know what is going to happen to his scores as a result of giving disingenuous responses,—even the authors cannot tell him except in a very few instances,—there seems to be little possibility of a systematically successful falsification of response on the free association test. In fact, the authors consider it the most nearly coach-proof test of all the psychological tests which involve conscious intelligent responses.

If it is possible to give a free association test, score it several ways for several character traits, and obtain reliable trait measures, a tool is made available for the study of a socially important and far-reaching type of phenomenon. The authors do not claim as yet to have accomplished this result, but the validity and reliability

TESTS AND MEASUREMENTS

of the trait measures thus far secured give promise that the problem is soluble in a practical sense.

In skeleton outline, the study may be divided into certain steps upon a chronological basis: the authors, familiar with the attempts at character measurement and the uses of the free association experiment in the study of abnormal psychoses and of normal interests, considered it a promising instrument to adapt to the measurement of character. The steps in the process appeared to be as follows:

- A. The selection of an appropriate list of stimulus words.
- B. The giving of these stimulus words one at a time to subjects, securing first responses that come to mind, thus catching native or deep-rooted tendencies to associate.
- C. The securing for these same subjects of trait ratings based upon judgments of associates, fellow students, and teachers, which ratings constitute entirely independent measurements of the traits it is desired to measure by means of the test.
- D. A weighting or attachment of appropriate credit to each response given by these subjects for each of the traits rated, thus providing a scoring scheme or means of grading the responses of a new subject to the same stimulus words.
- E. A checking of the validity and reliability of the instrument, including its attached scoring scheme, by comparison of the trait scores earned by the members of a new group with entirely independent trait ratings.

The steps just mentioned were undertaken and completed under the auspices of the Commission on Investigation of the Social Studies in the Schools. They resulted in an instrument with reliability for each of eight character traits in the neighborhood of .35. This reliability being inadequate for individual measurement, the authors were aided by a grant from the Carnegie Foundation for the Advancement of Teaching given to Harvard University by means of which they have been able to investigate the adequacy of the instrument as a means of measuring the mean standing of class groups in the same eight traits.

CHARACTER TRAITS AND CRITERION MEASURES

This study has revealed certain weaknesses in procedure which should be remedied in the future, and, in the opinion of the authors, confirms their earlier judgment that the instrument is a valid one for the study of character traits in groups. There remains the essential process of refining and modifying it still further, so as to make it a trustworthy instrument for the measurement of individuals. Suggestions to this end will be made later.

SECTION II

SELECTION OF CHARACTER TRAITS AND CRITERION MEASURES

Critical examination of numerous discussions of character, morals, ideals, citizenship, and the like indicated that writers have frequently attached very different meanings to the same trait name, as well as different trait names to the same type of social behavior. Statistical tabulations showing the relative frequency with which each trait appears in the literature were therefore not considered entirely appropriate as final determinants of the characteristics to be employed in this experiment with the free association test. Additional criteria were accordingly adopted and used in reducing the number of traits to eight, which was thought to be as large a number as could be handled in such a study.

In the first place, the effort was made to select only those traits which are likely to have real and immediate significance in making adjustments to social institutions and to one's fellows. Such characteristics as an attitude of emotional detachment to the conflicts reported in history may be of great value to a young scholar in the library, but the connection between this attitude and the making of happy adjustments with his fellows and with his immediate social environment is not so clear and certain. Unless the connection between the presence of the trait and the development of desirable personal and social relationships was fairly obvious, a trait was not included in this study.

A second standard used in the selection of characteristics was the degree to which a trait differed from the other traits being considered. Persistence, for example, is an often-recommended char-

TESTS AND MEASUREMENTS

acter trait, but it is closely related to the trait called "school drive," which had been chosen for other reasons. While the inter-correlations of the measures developed in this study indicate that there are common elements in the traits selected, a serious effort was made to choose for investigation traits that seemed to be as distinct or discrete as possible.

A third criterion of selection grew out of the method to be used for obtaining independent measures of the traits selected. Since each pupil was to be judged by his fellows and by his teachers as to the degree to which he possessed each of the eight traits, it was desirable to select only those traits in which evidences of presence or absence could be readily observed and accurately estimated by teachers and pupils. It would have been wasteful of effort to have included a trait in which ratings by those who knew him best would be unreliable or lacking in validity.

Closely related to this was the fourth criterion, that only those traits should be chosen for experiment in which other useful techniques of measurement had already been developed or might, in the judgment of the authors, soon be developed. While the practice of validating one test by its correlations with another has definite limitations, it is certainly desirable in the development of any new testing technique to be able to check the results of the new method against the results obtained from independent measures of the same traits.

While the traits selected by the application of these criteria may not be the most important elements of social character, they were as good a selection as the authors could make for their study at the time the choices had to be made. It was assumed that, if the free association test technique could be made to reveal the relative strength or weakness of each of these eight traits, it would be possible later to develop scoring keys for revealing the strength of any other equally identifiable trait or attitude.

The eight traits chosen for the experiment are indicated herewith, each with brief characterizations of both its positive and its negative behavior patterns.

CHARACTER TRAITS AND CRITERION MEASURES

COURTESY

(Positive)

Considerate, courteous, polite, gracious, diplomatic, refined, well-bred, tactful, gentlemanly (ladylike), chivalrous, urbane.

(Negative)

Inconsiderate, discourteous, impolite, rude, unmannerly, undiplomatic, crude, ill-bred, tactless, ungentelemanly (unladylike), barbarous, boorish, rustic.

FAIR PLAY IN PLAYGROUND GAMES AND CONTESTS

(Positive)

Always plays by the rules, even when it means that his side may lose.
Insists on applying the same rules in the same way to both sides.
Never tries to cheat or to take an unfair advantage of the other side.

(Negative)

Does not play fair or follow the rules except when they favor his side.
Tries to apply the rules in one way for his own side and in another way for the other side.
Takes unfair advantage of the other side whenever he has a chance.

HONESTY IN SCHOOL WORK

(Positive)

Does not copy materials from other children or from books.
Does not cheat on tests and examinations.
Reports practice-test scores correctly.
Does all his own work exactly as he was told to do it.

(Negative)

Hands in, as his own, materials copied from books and from other pupils.
Cheats on tests and examinations.
Lies about his scores on tests.
Gets help in doing assignments that were to be done by himself.

LOYALTY TO HIS FELLOWS

(Positive)

Is always loyal to his chosen leaders and playmates.
Will sacrifice his own pleasure for the good of his gang.
Never "tells" on a member of his gang.
Stands by his friends "through thick and thin."

(Negative)

Is not loyal to his leaders and playmates.
Will sacrifice his gang for his own pleasure.
"Tells" on his playmates if he thinks it will help him.
Deserts his friends when they are in trouble.

MASTERY

(Positive)

Masterful, commanding, dominant, authoritative, domineering, active, rough, forceful.

(Negative)

Submissive, obedient, compliant, deferential, yielding, easily led, passive, gentle, mild.

TESTS AND MEASUREMENTS

POISE

(Positive)

Composed, calm, confident, poised, self-controlled, emotionally stable, well-balanced, judicious, prudent, discreet, consistent, careful, deliberate, thoughtful.

(Negative)

Nervous, anxious, agitated, flighty, easily-excited, emotionally unstable, eccentric, injudicious, imprudent, indiscreet, inconsistent, reckless, hasty, thoughtless.

REGARD FOR PROPERTY RIGHTS

(Positive)

Always recognizes an owner's right to control and to protect his own property.

Asks permission before he borrows anything; returns things promptly and in good condition.

Avoids injury to property of others, but tries to make good any injury done accidentally.

Tries hard to find the owner in order to return anything he has found.

(Negative)

Takes things he wants, regardless of who owns them.

Forgets to return and does not take care of things he has borrowed.

Deliberately defaces and injures the property of the school and of others.

Runs away and denies responsibility when he has injured someone's property.

Keeps anything he finds, even when he knows to whom it belongs.

SCHOOL DRIVE

(Positive)

Tries to learn all his lessons perfectly.

Gets to work immediately without any complaints.

Does not give up when a lesson is long and hard.

Tries to do well everything the teacher asks him to do.

(Negative)

Never does an assignment well if he can get out of it.

Complains and objects to assignments.

Gives up trying to get a lesson as soon as he can find an excuse.

Frequently prevents others from doing good work.

The above words and phrases descriptive of each trait and its opposite were used in obtaining ratings by teachers and supervisors of the pupils in each class examined. Each teacher was asked to indicate for each trait those pupils in the class who were outstandingly high or low in the quality. The sheets on which these different ratings of a class were obtained were identical except for the titles and descriptions of the traits. Facing this page is a reproduction of the 8½"x11" sheet on which these ratings were made. The mimeographed instructions to the teacher making the ratings were as follows:

DIRECTIONS FOR RATINGS OF PUPILS BY TEACHERS

1. Do not attempt to consider more than one trait at a time. In connection with each of the eight characteristics, take at least two days in which to observe closely the behavior of the pupils.

TEACHER'S RATINGS OF PUPILS

Room No. _____ Grade _____ Teacher _____ School _____

Total enrollment of class _____ **Boys** _____ **Girls** _____ **Ratings by** _____

Please read carefully the descriptions printed below of the two extreme degrees of the trait; and then observe the pupils, with that trait in mind, for at least two days before recording your judgments. The scientific worth of this study depends directly upon the validity of your ratings of the pupils.

Please have before you when making your estimates a complete alphabetical list of all the pupils in the class. If two or more pupils seem to you to be equal in this trait, please indicate equality by enclosing their names in brackets.

LOYALTY TO HIS FELLOWS

(Positive)

Is always loyal to his chosen leaders and playmates.
Will sacrifice his own pleasure for the good of his gang.
Never "tells" on a member of his gang.
Stands by his friends "through thick and thin."

Please list at the top of this column the name of the pupil who has the greatest amount of the positive trait. List second the name of the pupil who has the next-largest amount, and so on.

(A moderate amount of the negative trait)

Please list at the bottom of this column the name of the pupil who has the greatest amount of the negative trait described below. Just above his name write the name of the one who has the next-largest amount, and so on.

(Negative)

Is not loyal to his leaders and playmates.
Will sacrifice his gang for his own pleasure.
"Tells" on his playmates if he thinks it will help
him.
Deserts his friends when they are in trouble.

(A moderate amount of the positive trait)

School Room No. Grade Teacher

WHO IS IT?

Every class has in it a few boys and girls who in some way or other are not just like the others. This is a test to see how well these unusual pupils are known to their classmates.

You need not sign your name to this test unless you wish to do so. Your answers will be used in a statistical study which is being made at a university and will probably never be seen by your teacher or principal. You may write your own name after any question to which you think it is the best answer. If you think no member of your class is like the person described in a question, skip that question and go on to the next. Do not forget to write both the first and last names of any classmates who are like the persons in the questions.

1. Who is it that always stands up for his (or her) companions, and never tells on a friend?

2. Who never gets excited, but is always calm and careful?

3. Who is it that always tries to boss everyone else and to run everything?

4. Who never tries to cheat or to take an unfair advantage of the other side when playing games?

5. Who is it that doesn't even try to get the lessons your teacher assigns, unless made to do it?

6. Who is most often rude and impolite to people?

7. Who is most nervous and gets excited most easily?

8. Who is it that takes other people's things without permission and tries to keep them?

(over)

9. Who is always polite and considerate of other people's feelings?

10. Who is it that never cheats or does anything that is unfair when taking examinations or getting lessons?

11. Who is always ready to go anywhere or do anything that anyone suggests?

12. Who is it that tells on playmates, and forgets friends when they are in trouble?

13. Who always tries to find the owner in order to return anything that has been found?

14. Who is it that does not play according to the rules and often takes an unfair advantage of the other side?

15. Who tries to do perfectly everything the teacher asks or assigns?

16. Who marks on the walls and breaks or injures property?

17. Who gets someone else to do his (or her) school work or copies answers instead of working them out alone?

18. Who always asks permission before borrowing anything, and always returns things promptly and in good condition?

YOU NEED NOT SIGN YOUR NAME TO THIS TEST.

CHARACTER TRAITS AND CRITERION MEASURES

2. Have a complete alphabetical list of the names of members of the class before you in making each rating.
3. Rate each pupil as accurately as possible for the trait described on the blank. Do not allow your general opinion of the pupil or your recorded judgment regarding him in some other characteristic to influence you.
4. It is not necessary to write on the blank the names of those pupils who have only an average or indifferent amount of the characteristic named. Record only the names of that third of the class having the greatest degree of the positive trait and the names of that third having the greatest degree of the negative trait. The names of those pupils who have the largest amount of the positive characteristic should be entered at the top of the column on the left, and the names of those having the largest amount of the negative trait at the bottom of the right-hand column.
5. As soon as you have rated a given characteristic, put away your record sheet and give your full attention to the next characteristic to be rated. Do not allow your record in one trait to influence your judgment in rating another trait.

ORDER OF RATINGS

1. Courtesy
2. Loyalty to His Fellows
3. Poise
4. Fair Play in Playground Games
5. School Drive
6. Regard for Property Rights
7. Honesty in School Work
8. Mastery

Two teachers who knew the members of the class well, usually their regular home-room teacher for the current year and their regular teacher during the previous year, made out independent rating sheets for each of the eight traits. Occasionally the second rating of a class was made by the principal, the general supervisor, or the teacher of some special subject. Having two independent ratings of the members of each class made possible the calculation of the reliability of these ratings. (See Tables VI and IX, pages 412, 413.)

Still another independent series of ratings of the members of each class on approximately the same eight traits was obtained by having each pupil fill out anonymously a "Who Is It?" blank. This blank and the directions for administering it are here reproduced.

TESTS AND MEASUREMENTS

WHO IS IT?

Directions for Administering

The chief points to be sought for in administering this test are (1) to secure independence of judgment from each pupil, and (2) to prevent pupils from being influenced by other considerations than the actual behavior of their classmates. Independence of judgment can be obtained by insisting that there shall be no communication or looking at what anyone else has written. Honesty of judgment is more difficult to obtain, because many of the pupils who have the keenest judgment have also the strongest sense of loyalty to their fellows. In case any hesitancy upon the part of the class is shown in writing names, an examiner should make it perfectly clear to the pupils that neither he, the teacher, nor the principal will look at the answers to this test, and that no member of the class will gain any advantage or disadvantage from the answers that are written to these questions.

See that each pupil is provided with a pen or pencil, and with a desk or table on which he can write comfortably. Hold up a copy of the test and make the following explanations:

"This exercise is different from the usual school exercise. You need not sign your names to your papers, but you are to give the best answers you can to each question. Do not read the questions until I tell you to do so, but fill in at once the name of the school, your class or grade, and the teacher's name." Distribute the papers, one to each pupil, with the title page up, and then read aloud the directions printed on the test blank:

"Every class has in it a few boys and girls who in some way or other are not just like the others. This is a test to see how well these unusual pupils are known to their classmates.

"You need not sign your name to this test unless you wish to do so. Your answers will be used in a statistical study which is being made at a university and will probably never be seen by your teacher or principal. You may write your own name after any question to which you think it is the best answer. If you think no member of your class is like the person described in a question, skip that question and go on to the next. Do not forget to write both the first and last names of any classmates who are like the persons in the questions.

"Write the names of two of your classmates after each question, if there are two who are like the person mentioned in the question. No pupil will gain any advantage or disadvantage because you write his name. I shall have you put your paper into this large envelope yourself," (hold up the container) "and then I shall send it away to a man who does not know you or your classmates. He will not report what you have said. Please be perfectly honest. Writing something that is not true would not help or hurt any of your classmates, but it might spoil a very important investigation which is being made at the university, and it would certainly not do you or anyone else any good. Answer each question as well as you can. Are there any questions?"

WORD ASSOCIATION TEST

In answering questions, use where possible the language of the instructions already given. Allow the pupils as much time as necessary to answer the questions.

In taking up the papers, have the child himself insert his paper in the envelope or do it yourself in his presence. Do not in the presence of the children take from the envelope any papers which the children have written, and do not allow the teacher to do so. If the children are to give us their honest judgments they must feel that they are being fairly treated.

By combining the ratings of the two teachers with the returns given by the pupils themselves on the "Who Is It?" blank, all members of each class were identified for each of the eight characteristics. The weight to be assigned any particular response to a stimulus word in the free association test was later determined statistically, as explained in the next section of this report, by considering the difference in the frequency with which that response appeared among the pupils who had been rated high and among the pupils who had been rated low in the trait.

The schools and classes in which these tests were administered were selected in terms of the following standards:—(1) The pupils must be normal representatives of the total school population who have not had any special training in social attitudes and who do not present any highly unusual behavior problems. (2) There must be at least two teachers or supervisors who know the pupils well enough to give useful ratings of their behavior patterns and attitudes. (3) There must be at least one teacher or supervisor who understands the importance of uniformity in the administration of the tests and who will take responsibility for seeing that they are given according to the directions. (4) City, village, and country schools should all be included in order to avoid the influence of any special type of environment. The village and country schools used were in North Carolina, while the city schools were in Washington, D. C.

ADMINISTRATION OF WORD ASSOCIATION TEST

Previous investigators had found that many persons, when allowed to write response words without supervision or control, tended to develop stereotyped modes of reaction. The character

of the examples used in explaining the test had been recognized as influential in determining the type of reaction adopted. If, for example, the response word given in an illustration was an opposite of the stimulus word, there was thereafter a tendency among those being examined to respond to each stimulus word by writing its opposite. Similarly, an example in which the response word was a part or specific instance of the thing named by the stimulus word tended to fix the whole-part relationship as the pattern to be followed in subsequent responses.

Several devices were tried by the present authors in an effort to prevent the development of such uniform types of response, which could obviously reveal little regarding the subject's personal attitudes or individual character. It was thought that it would probably be advantageous to use in the directions a number of different illustrations, each representing a distinct type of association, but all being of the character that might have significance. Among the responses to be avoided as types, though not as occasional indicators of associative process, were the "clang" reaction, in which the response is a word that sounds like or rhymes with the stimulus word, the "opposite" reaction, and the "whole-part" response mentioned above, the response which consists in giving a synonym or definition of the stimulus word, and the "predicate" response, in which one indicates something which is done by or with the thing for which the stimulus word stands.

A further attempt to prevent the development of such stereotyped responses was made by providing printed lists of response words in which the pupil being examined was to check the first response that had come to his mind. It was also hoped that the provision of these check lists might possibly reduce the amount of time required to secure responses and thus make possible an increased number of stimulus words in a single testing period. The printed lists were made up of those responses which were known to be most frequently given to each stimulus word, except that "opposites," "clang" responses, and the like, were not always

WORD ASSOCIATION TEST

included. Any child who actually thought of one of these stereotyped responses which had been omitted would have to write out his response word. The purpose was not to rule out any type of reaction, but to encourage the more distinctly individual responses.

Experimentation with mimeographed and printed pads, on each page of which one stimulus word and twenty appropriate response words appeared, indicated that many pupils were inclined to check the response words appearing at the tops of the two columns rather than those at lower points on the page. In other words, pupils were marking the first response that came to their attention from the printed page rather than the response which came from personal modes of mental reaction. To prevent their continuance in this practice, pupils were asked to enter in a half-inch square at the bottom of each page the first letter in the first word of which they had thought when the "key-word," *i. e.* the stimulus word, for the next page was pronounced. After this initial letter had been recorded, they were asked to turn the page and check or write the entire word. Even this procedure failed to prevent a few pupils from depending on the printed list for suggestions, but the number of such pupils was greatly reduced by the device.

In working on the administration of the association test, one of the important considerations in the minds of the authors was to obtain free responses to as many different words as possible, in order to secure maximum reliability in the resulting indexes to the individual's habitual thought patterns. With twenty response words printed on each page, many pupils in each class still found it necessary to write their responses. There was some evidence that twenty response words printed alphabetically in two columns required an undue amount of time on the part of certain pupils in the mere process of determining whether their own responses were among the twenty. The number of printed response words was therefore reduced to fourteen on each page, the stimulus word was omitted from the page, and the organization of the test pad or book was simplified. Each of the 200 pages following the first three, which were explanatory examples, was similar to page 49,

TESTS AND MEASUREMENTS

which is here reproduced, and for which the stimulus word was **SILK**.

.....clothgoodssatinwool
.....cottonhosesoftworm
.....dressJapanstockings	
.....fabricmaterialthread	



49

Stimulus words were selected from those previously used by other investigators, in order that those which frequently stimulate a single response, a narrow range of responses, or no response might be omitted, and in order to be able to use the findings of these investigators in choosing the most frequent responses to be printed on the test page devoted to each stimulus word. The Woodrow-Lowell¹ tabulations of children's responses to the Kent-Rosanoff words, Johnson O'Connor's² tabulations of adult responses to the same words, Jennie B. Wyman's³ tabulations of children's responses, and P. E. Schellenberg's⁴ tabulations of the responses of five hundred college students were the chief sources from which the fourteen response words to each of the two hundred stimulus words were selected. An additional hundred stimulus words, each with twenty of its most common responses, were chosen at the same time, but these were not used in this study on account of the great length of time that would be required to secure responses to each of three hundred stimulus words. Approximately an hour and three quarters was required to secure responses from a junior-high-school class to two hundred words when given by means of the printed test books according to the following instructions:

¹Woodrow, Herbert, and Lowell, Frances. *Children's Association Frequency Tables*. Lancaster, P. The Psychological Review Company, 1916.

²O'Connor, Johnson. *Born That Way*. Baltimore, Williams and Wilkins Co., 1928.

³*On the Influence of Interest on Relative Success*. A Dissertation on file in the library at Stanford University.

⁴A Dissertation on file in the library at the University of Minnesota.

WORD ASSOCIATION TEST

NEXT-WORD EXERCISES

INSTRUCTIONS FOR ADMINISTERING

See that each pupil is comfortably seated at a desk or writing table and is provided with a pen or pencil which is in good working condition. Hold up a copy of the booklet and tell the pupils that you are going to give them an unusual but very interesting "word exercise." Direct them to fill in at once their names and the other facts called for on the front cover of the booklet. Ask them not to open the books or to look inside until they are told to do so. Distribute the booklets, face up, and see that the names and other data are entered quickly and correctly.

As soon as the blanks on the cover page have been filled, read slowly and distinctly the following explanation:—

"This is an exercise in which you are to tell the next word you think of after each word I pronounce. I shall say a word, and then you are to think of some other word which it suggests to you. Any word will do, if it is the first word that comes to your mind after I pronounce my word. The first letter of the word you think of is to be written in a little square before you turn the page, and then the entire word is to be written on a dotted line after you turn the page.

"Now we are ready to try a word. As soon as you have thought of your word and written its first letter in the little square, hold up your hand. Never turn a page until you are told to do so.

"Do you all see square No. 1, near the bottom of the page?" Pause until everyone has found it. In the lower elementary grades it will be helpful to suggest that each child put his finger on the number just below the square.

"In square No. 1 write the first letter of the word you think of when I say the word 'LAMP.'" Pronounce it distinctly and then spell it out, "L-A-M-P. Write the first letter of the word you thought of, and then hold up your hand." Pause until all but two or three hands are up. Observe which pupils fail to get their hands up, so that extra help may be given them after the other pupils have turned the page.

"Now turn to page 2. Write on the dotted line the word you thought of. Hold up your hand as soon as you have written your word." Pause until 90% of the hands have been raised, and then ask, "Do all of you understand what to do when I tell you the next word? First, you think of another word. Second, you write the first letter of your word in the square. Third, when I tell you to turn the page, you turn to the next page and write all of your word on the dotted line. Are there any questions?" Answer briefly any questions that may be asked.

"Do all of you see square No. 2? Write in square No. 2 the first letter of the first word you think of when I say the word 'RICH, R-I-C-H.' Hold up your hand as soon as you have done it." Pause until about 90% of the hands are up, and then say, "Turn to page 3 and write the word you thought of. Hold up your hand as soon as you have written it." Pause until all but one or two hands are up.

"Now look at square No. 3. Write in it the first letter of the word you think of when I say the word 'RIDE.' Then hold up your hand." Pause. "Now turn the page and write the whole word on the dotted line." Pause until all but two or three have finished.

TESTS AND MEASUREMENTS

"Now read to yourselves, while I read aloud what it says just above square No. 4. 'On each of the next pages you will find 14 words, printed alphabetically in two columns. On each page, if the word you thought of is in the list, put a check mark in front of it. If the word you thought of is not in the list, write it on the dotted line.' Pause.

"Do all of you know how to look for a word in an alphabetical list? Suppose the word you think of begins with the letter B. Where would you look for it in an alphabetical list?" Pause for an answer.

"Yes, that's right! You would look among the words that begin with B, near the top of the first column. Now suppose the word you think of begins with the letter Y. Where would you expect to find it?" Pause for replies. "Yes. You would look among the words beginning with Y, near the end of the second list. When I tell you to turn a page, do not look at all the printed words. Look only at the words which begin with the same letter as your word. If the word you thought of is not where it ought to be in the list, it is not there at all, and you will have to write it, as you have written the first three words. If you find your word in the printed list, make a little check mark in front of it, instead of writing it.

"Now look at square No. 4. In square No. 4, write the first letter of the word you think of when I say the word, 'AGE.' Then hold up your hand." Pause. "Now turn to page 5 and see if the word you thought of is in the printed list. If so, make a check mark in front of it. If it is not there, write it on the dotted line." Pause. "Does anyone wish to ask a question?" Answer any questions. Be sure that the pupils understand the procedure, and then continue as rapidly as possible, using the following directions for the next half dozen pages:

"In square No. ... write the first letter of the word you think of when I say the word " Pause. "Now turn to page ... and check your word, or write it." As soon as the routine is established, abbreviate the directions to the following:

"The word for square No. ... is " Pause. "Turn to page " Do not neglect to give the number of the page and square each time, because it is very easy for the pupil to turn more than one page, especially when the last pages have been reached.

Two or more class periods will be required for completing the test. Do not try to give the entire test at one sitting. Allow at least one rest period during the test.

When the initial letter has been written in square No. 101, have the pupils turn their books upside down to find page 102. See that everyone has found the place before pronouncing further words.

STIMULUS WORDS FOR NEXT-WORD EXERCISES

In pronouncing these words, bear in mind that the word listed for page 5 is pronounced while the pupils are still looking at page 4; that the pupils write the first letter of their response word in the square on page 4; and that they check or write their entire response word on page 5.

WORD ASSOCIATION TEST

Page No.	word	Page No.	word	Page No.	word	Page No.	word
1	51	whistle	102	wait	152	white
2	lamp	52	cottage	103	wild	153	water
3	rich	53	eagle	104	over	154	silly
4	ride	54	hurt	105	note	155	sad
5	age	55	kitchen	106	difficult	156	cold
6	dress	56	pain	107	desire	157	child
7	stranger	57	platform	108	sickness	158	hit
8	sundown	58	light	109	shock	159	hate
9	night	59	lion	110	black	160	vacation
10	ocean	60	crowd	111	band	161	try
11	school	61	dance	112	seed	162	butter
12	music	62	tire	113	once	163	bread
13	needle	63	train	114	mountain	164	slow
14	restless	64	friendly	115	milk	165	sky
15	river	65	habit	116	evening	166	castle
16	poor	66	smooth	117	eating	167	camp
17	prison	67	soldier	118	thief	168	mischief
18	rain	68	king	119	swift	169	love
19	range	69	learn	120	home	170	sweet
20	mother	70	flag	121	admire	171	street
21	name	71	form	122	companion	172	cotton
22	church	72	join	123	cave	173	correct
23	citizen	73	joy	124	pure	174	man
24	pleasure	74	fair	125	sorrow	175	make
25	prefer	75	fault	126	sing	176	stem
26	square	76	country	127	dark	177	spider
27	stage	77	deep	128	current	178	protect
28	loud	78	snake	129	fairy	179	power
29	marble	79	star	130	exercise	180	coat
30	command	80	queer	131	journey	181	choose
31	cow	81	college	132	island	182	must
32	stove	82	control	133	foot	183	mouth
33	summer	83	fond	134	fire	184	red
34	listen	84	honor	135	laugh	185	quiet
35	luck	85	table	136	sour	186	priest
36	boy	86	easy	137	soft	187	pray
37	captain	87	enjoy	138	hair	188	rough
38	simple	88	master	139	give	189	ring
39	sleep	89	moon	140	travel	190	need
40	blue	90	new	141	tobacco	191	nature
41	butterfly	91	rose	142	date	192	science
42	trip	92	aim	143	cut	193	salt
43	useful	93	bat	144	long	194	nut
44	handle	94	sheep	145	like	195	nice
45	hay	95	short	146	playing	196	Sunday
46	cheese	96	delighted	147	people	197	stomach
47	city	97	diamond	148	lazy	198	dream
48	rush	98	narrow	149	important	199	beat
49	silk	99	old	150	earth	200	dog
50	watch	100	walk	151	doctor	201	dislike
		101	young				

TESTS AND MEASUREMENTS

SECTION III

THE DEVELOPMENT OF THE SCHEME FOR SCORING THE ASSOCIATION TEST

Though it might be within the power of some peculiarly perspicacious individual to look at each response word and to surmise the nature of the probable associative process so accurately as to enable him to assign scores for (1) Courtesy, (2) Fair Play in Games, (3) Honesty in School Work, (4) Loyalty to his Fellows, (5) Mastery, (6) Poise, (7) Regard for Property Rights, and (8) School Drive (each as defined in Section II), the writers credited themselves with no such ability and have resorted to a strictly objective method to determine the credits that should be attached to the different responses.

As has already been explained the judgments of teachers and fellow pupils were utilized to obtain criterion scores for each of the eight traits. These were combined for each trait and so expressed that each grade group studied had criterion scores of mean value zero and standard deviation equal to 1.00. Consider the first trait only and some particular response R to a particular stimulus S. If ten pupils gave this response R (out of the total of 723) we may list the criterion scores for these ten. If, for the first trait, they are as follows, .8, .6, -.4, 1.0, .7, .0, .3, .0, 1.5, .5, the sum equals 5.0 and the average is .5. Thus, if a child responds with the word R to the stimulus S, the most likely estimate of his trait 1 criterion score is .5. We accordingly attach a weight (slightly modified as will be explained) of .5 to this response word. This constitutes the "score" for this response R to this stimulus S for this first trait. Similar scores are computed for all other responses.

One way of investigating the merit of such a scoring scheme is to determine what would result if all response words were the result of chance so far as the trait is concerned. Clearly the ten individuals giving this response would not all have criterion scores of zero, for if that were the case the response R is definitely indicative of mean trait 1 standing. If the response is in truth non-

SCHEME FOR SCORING THE ASSOCIATION TEST

significant we will expect the ten scores to yield a distribution which is a random distribution from a parent population (of 723) whose mean is zero and standard deviation 1.0. Only in so far as the obtained distribution differs from this normal expectation should we believe that there is evidence of actual correlation between response R and trait 1. Thus a modification of such a value as the .5 just given is called for. This may be a very radical modification if the number in the class giving the response R is small.

For example, if a large number of pupils give the response i , then the mean value of their criterion scores $\bar{z}_{0.i}$ calls for no modification, it being the proper value to assign to this response when some new pupil gives it. If but two or three give the response j , then $\bar{z}_{0.j}$ is not as trustworthy a score when attached to the response of some new pupil as is $\bar{z}_{0.i}$. In other words, $\bar{z}_{0.j}$ should be discounted. We herewith give the determination of the appropriate discounting factors.

Let responses be designated by $a, b, c, \dots i, j, \dots$

Let number of i responses = n_i

Let mean z_0 , i.e. criterion, scores of these = $\bar{z}_{0.i}$ and let the n_i such scores be designated by $z_{0.i1}, z_{0.i2}, \dots, z_{0.in_i}$.

Let the z_0 score for an added i response, i. e. one not made by any of the n_i subjects used in determining the weight, be designated $z_{0.i}$ and if necessary by $z_{0.i1}, z_{0.i2}, \dots$

Remembering that z_0 are standard measures, so that $\Sigma z_0 = 0$ and $\frac{1}{N} \Sigma z_0^2 = 1.00$, let us determine the variance of estimate when $\bar{z}_{0.i}$ is taken as evidence of $z_{0.i}$. Let $\bar{z}_{\infty.i}$ = true mean z_0 score for an infinite population giving response i , and let $\sigma^2_{0.i}$ = the variance of such scores from $\bar{z}_{\infty.i}$. Thus if $\bar{z}_{\infty.i}$ were taken as evidence of $z_{0.i}$ the variance error = $\sigma^2_{0.i}$. When $\bar{z}_{0.i}$, having a variance from $\bar{z}_{\infty.i}$ of $\frac{\sigma^2_{0.i}}{n_i}$ is taken in lieu of $\bar{z}_{\infty.i}$, the variance error is

$$\sigma^2_{0.i} + \frac{\sigma^2_{0.i}}{n_i}, \text{ or } \sigma^2_{0.i} \frac{n_i + 1}{n_i}.$$

TESTS AND MEASUREMENTS

The weight that should be given to $\bar{z}_{0,i}$ is inversely as the variance, i. e. it equals

$$\frac{n_i}{(n_i + 1) \sigma^2_{0,i}}$$

The labor involved in getting an estimate of each $\sigma^2_{0,i}$ for each of the 5000-odd responses is prohibitive. We must therefore

assume them equal and weight each $z_{0,i}$ as $\frac{n_i}{n_i + 1}$, as shown in the accompanying table.

n_i	1	2	3	4	5
Multiplier of $\bar{z}_{0,i}$.5000	.6667	.7500	.8000	.8333

6	7	8	9	10	11	12	13
.8571	.8750	.8889	.9000	.9091	.9167	.9231	.9286

14	15	16	17	18	19	20	21
.9333	.9375	.9412	.9444	.9474	.9500	.9524	.9545

22	23 25	26 28	29 31	32 34	35 37	38 40	41 43
.9565	.9600	.9643	.9677	.9706	.9730	.9750	.9767

44 46	47	48 52	53 57	58 62	63 67	68 69	70 79
.9783	.9792	.9804	.9821	.9836	.9848	.9856	.9868

80 89	90 99	100 on
.9883	.9895	1.0000

SCHEME FOR SCORING THE ASSOCIATION TEST

We thus see that if the credit attached to a certain response is consequent to a criterion score of but one person out of the total of 723, this score is multiplied by .5, that is, discounted 50 per cent; if consequent to criterion scores of two persons, it is multiplied by .67, that is, discounted 33 per cent; etc.

Though the credits and debits computed by means of these weighting factors may be the best possible, it still should be observed that they will not constitute very reliable scores if the numbers giving certain responses are small. For an initial sample of 723, yielding 28 response words,—more accurately, res classes,—the average number giving a certain response is 26. If this adequately indicated the reliability with which a score is attached to a response, the final outcome in terms of the reliability of the scoring scheme would be excellent. It, however, does not, for this mean 26 is probably derived from many small classes and a very few large classes. The reliabilities of scores computed from these small classes are ordinarily to be represented by frequencies from 1 to 10, and the reliability of the two or three large classes is represented by

$$\frac{n_i}{(n_i + 1) \sigma_{0.1}^2}$$

and not by

$$\frac{n_i}{(n_i + 1)}$$

Since the $\bar{z}_{0.1}$ means for these large classes are found by observation to differ but slightly from .0, there is evidence that for these classes $\sigma_{0.1}$ is larger than for the remaining classes. If $\sigma_{0.1} = \sigma_0$, the scores based upon these large classes would have no diagnostic value, for all such scores would be .0.

It thus appears that scores based upon large classes cannot be expected to have much diagnostic value, and that it is only as children reply with more or less unique response words that real character divergencies from the mean, if present, will be revealed. In short, the solution to this problem depends upon ability reliably

TESTS AND MEASUREMENTS

to attach credits and debits to small class responses. This can only be done if the initial sample is much larger than 723.

An alternative procedure which has been suggested is to use judgments of psychologists in attaching credits and debits. The possibility of this is admitted, depending upon the insight of the psychologists. To date, this procedure has not been followed and it must always be considered secondary to the method outlined in the preceding paragraphs, wherein judgments have been utilized in grading children for character traits, not in grading association responses. The former is the normal channel wherein judgments operate, not the latter.

SECTION IV

RESULTS OF THE STUDY TO JANUARY 1, 1932

Financial assistance from the Commission on Investigation of the Social Studies of the American Historical Association roughly sufficed to carry the study to January, 1932, and to the accomplishment of the following things: (1) the completion of the first administration of the test; (2) the determination for the largest of the three groups studied of the scoring scheme for this grade level; (3) the determination of the reliability coefficients of each of the 8 trait measures; and (4) the correlation of each with the corresponding 8 criterion scores. It became obvious at this point that the test was not sufficiently reliable to yield satisfactory individual measures.

To determine its value as an instrument for group measurement the Carnegie Foundation for the Advancement of Teaching, through the Graduate School of Education, Harvard University, provided funds for the continuation of the study. By January, 1932, the test had been given to 435 children of average school grade 5.0 (beginning of the fifth grade,—children were drawn from fourth, fifth, and sixth grades), to 800 children of average school grade 7.0 (children drawn from sixth, seventh, eighth, and ninth grades), and to 629 children of average school grade 9.0 (children drawn from sixth, seventh, eighth, ninth, tenth, eleventh, and twelfth grades).

RESULTS OF THE STUDY

It will be noticed that there is a certain amount of overlapping in the classes from group to group. The ninth graders in the group of 800 children of average grade 7.0 had Stanford Achievement Test scores below 78, or lacking such scores were of chronological age 16 or above, whereas the ninth graders in the group of 629 children of average grade 9.0 had Stanford Achievement Test scores of 79 or higher, or lacking such scores were 15 years of age or younger. The duller ninth graders were incorporated into the group of mean school grade 7.0 and the brighter into the group of mean school grade 9.0. Similar demarcations were made for the pupils of each other grade. Funds proved inadequate for working up the data for all groups, and data for the two smaller groups have been laid aside, though the association test responses are available, should it prove possible to study these at a later date.

The original seventh-grade group was divided into two parts: 723, which will be referred to as the standardization group, were used for the determination of the scoring scheme; and 67, referred to as the reserve group, were laid aside to be used after the scoring scheme derived from the 723 was available, to determine the reliability of each of the 8 measures and to determine the correlation of each with the respective criteria. The necessity for such a reserve group is obvious, for the scoring scheme for each trait was so built up as to make the correlation between the responses as scored and the criteria as given high. Clearly, having done this, we cannot then investigate the correlation with the criteria for these 723 cases and consider the result fair as a statement in general. In fact it would be spuriously high, but how much too high would be difficult to determine, and is unknown. Because of this, it is believed that such a correlation coefficient would have no value, and it has not even been calculated. Neither have any reliability coefficients been derived from this group of 723, as these would also be spuriously high.

The reserve group of 67 is too small to give satisfactory measures of reliability, but it was made small in order to retain as many cases as possible in the essential undertaking of determining the scoring scheme. Even the standardization group of 723 is not of

TESTS AND MEASUREMENTS

sufficient size to serve well its purpose. Table I presents the reliability coefficients found for this group of 67, using the split-test method. The responses to the 198 stimulus words were divided into 12ths as nearly as could be, the score on the odd 12ths constituting a one-half test score, and that on the even 12ths the other half test score. For each trait the correlation between these two was then stepped up by the Spearman-Brown formula, giving the reliability coefficients for the entire test of 198 words.

TABLE I
RELIABILITY COEFFICIENTS FOR RESERVE GROUP

Courtesy.....	.46
School Drive.....	.72
Regard for Property Rights.....	.74
Poise.....	.64
Mastery.....	.44
Loyalty.....	.56
Honesty.....	.52
Fair Play.....	.62

It was upon the basis of these reliabilities that it was concluded that none of the measures was accurate enough to be used for individual measurement, but that probably some of them would be of service in group measurement. As reported later, reliabilities as high as these have not been found to maintain when delinquent boys are tested. Some of the reasons therefor have come to light, but a remedying of the test in the light thereof has not been possible. The correlations with the criteria for this reserve group are presented in Table II.

TABLE II
CORRELATIONS BETWEEN ASSOCIATION TEST AND JUDGMENTS
OF TRAITS FOR RESERVE GROUP

Courtesy.....	.22
School Drive.....	.26
Regard for Property Rights.....	.17
Poise.....	.23
Mastery.....	.02
Loyalty.....	.10
Honesty.....	.28
Fair Play.....	.15

RESULTS OF THE STUDY

The low value for mastery might be accounted for on the basis of a halo effect in those whose judgments were incorporated into the criterion scores, for mastery is generally negatively correlated with the other traits. The generally low values of these validity coefficients are only in small part due to the unreliabilities of the criteria, and in view of the size of the reliability coefficients, they seem explicable only on the basis of chance.

When calculating the reliability coefficients it was observed that the reliability of the first 16 words with the second 16 was higher than that of the next to the last 16 with the last 16. It was thought that this might be due to pupils' getting lazy and developing a tendency to check some printed word, rather than to write in the first words that came to mind. To check up upon this point, a new group of 190 seventh-grade pupils in the Bryant Junior High School of Minneapolis, Minnesota, were given the test without booklets. The reliabilities found are as follows:

TABLE III
RELIABILITY COEFFICIENTS FOR 190 SEVENTH GRADERS

Courtesy.....	.44
School Drive.....	.57
Regard for Property Rights.....	.48
Poise.....	.45
Mastery.....	.00
Loyalty.....	.31
Honesty.....	.57
Fair Play.....	.56

These, in fact, are a little lower rather than higher than the reliabilities from the reserve group. Also the correlations with criteria for this group of 190 are low.

TABLE IV
CORRELATIONS BETWEEN ASSOCIATION TEST AND JUDGMENTS
OF TRAITS FOR GROUP OF 190 SEVENTH GRADERS

Courtesy.....	.18
School Drive.....	.16
Regard for Property Rights.....	.12
Poise.....	.00
Mastery.....	.15
Loyalty.....	-.10
Honesty.....	.12
Fair Play.....	.07

TESTS AND MEASUREMENTS

Criteria scores for these pupils were not based on as many judgments per pupil as for the reserve group, and accordingly they may be thought of as of lower reliability. Even so, the showing cannot be considered favorable.

For the original sample of 723, the reserve group of 67, and the seventh grade group of 190, the subjects dealt with may be assumed, so far as character is concerned, to have constituted a normal population. It remained to be seen whether the test would differentiate greater differences in these character traits than were ordinarily to be found in normal school groups. One would expect that delinquent groups, Ethical Culture groups, and groups having been given special character training would be differentiated one from another, even by means of a test of reliability not exceeding .3. It seemed more important to determine this, using the available measuring instrument, than to attempt immediately to refine it by extending and improving the scoring scheme.

The indicated lines for improvement were, (1) to modify the form of administration, eliminating all printed responses which by their presence might alter the pupil's reaction; (2) to present words orally and by means of large-type flash cards, so as to reduce to a minimum any uncertainty or annoyance due to difficulty of comprehension of the stimulus word; (3) to extend the scoring scheme to include less frequent words than those thus far provided for, but (4) this could be done satisfactorily only by greatly extending the number of subjects used in building up the scoring scheme; (5) to re-determine the traits to be measured, securing a number, probably less than eight, which are more independent of each other than the present eight; (6) to secure criterion scores more uncontaminated by halo effect. Changes (1) and (2) were made in giving the test to all delinquent and specially trained groups, and we believe with good results, for difficulties connected with administration were lessened. Though it was reasonable to expect that much improvement in the test would follow the doing of (3), (4), (5), and (6), it nevertheless seemed best first to keep the instrument as developed and to try it out on more varied groups than those to which it was originally applied, with the assumption that

RESULTS OF THE STUDY

group measures only would be of sufficient reliability to lead to important conclusions.

SECTION V

ADMINISTRATION OF TESTS IN 1932

The use of test pads or books, having on each page fourteen printed responses to the corresponding stimulus word, did not shorten the time required to administer the association test. On the contrary, it appeared that looking through the printed list of responses actually added to the time required of those pupils who failed to find their own responses in the list. A few pupils became confused when asked to write the first letter of their response on one page and the entire word on the next page. Occasionally a child lost time by looking for his response on the wrong page. The high cost of printing, binding, and transporting the two-hundred page pads contributed, also, to the decision to abandon their use in favor of the simpler procedure of having each pupil write his response words in a four-page folder.

Another difficulty which had come to light was that pupils sometimes failed to hear the stimulus word correctly. It seemed that it might be desirable to present the stimulus word to the eye and to the ear simultaneously, in order to avoid any possibility of a misunderstanding. Flash cards were therefore prepared, each about $14 \times 3\frac{1}{4}$ inches in size, with the stimulus words in standard lower case block letters that could easily be read at a distance of forty feet, the height of the humdrum letters, i. e. neither non-ascenders or descenders, being $15/16$ inches. The administration of the test was then changed to have the examiner hold up before the class a flash card containing the stimulus word while he pronounced it twice distinctly. The number of the stimulus word appeared on the flash card and was spoken by the examiner, so there was little chance of the pupil failing to understand which blank on the test leaflet was to be used in recording his response word.

As a device for directing the pupil's attention away from stereotyped responses, the checking of response words was

TESTS AND MEASUREMENTS
(Page One of Test)
SOCIAL SCIENCE
WORD ASSOCIATION TEST

Name _____ Grade or Class _____
Age at last birthday _____ years Teacher _____
School _____ City _____

DIRECTIONS

What is the next word you think of after each word is spoken? If it is in the list, put a check mark (✓) in front of it. If it is not in the list, write it on the dotted line.

Word A	Word B	Word C
_____ busy	_____ odd	_____ above
_____ circus	_____ among	_____ actor
_____ city	_____ break	_____ astronomy
_____ few	_____ cake	_____ comet
_____ fire	_____ equal	_____ dim
_____ fun	_____ even	_____ dreams
_____ gang	_____ family	_____ dust
_____ group	_____ give	_____ earth
_____ hot	_____ great	_____ falling
_____ jam	_____ half	_____ guide
_____ many	_____ it	_____ heaven
_____ men	_____ money	_____ high
_____ noise	_____ numbers	_____ meteor
_____ number	_____ part	_____ morning
_____ push	_____ pieces	_____ north
_____ ran	_____ river	_____ player
_____ school	_____ split	_____ shining
_____ street	_____ spoils	_____ sparkling
_____ theater	_____ time	_____ sun
_____ trouble	_____ two	_____ twinkle

ADMINISTRATION OF TESTS IN 1932

(Page Two of Test)

DIRECTIONS

What is the next word you think of after each word is spoken? Write the word you think of on the dotted line after its number.

1	26.....	51.....
2	27.....	52.....
Etc. to No. 25.	Etc. to No. 50.	Etc. to No. 75.

(Directions for Test)

SOCIAL SCIENCE WORD ASSOCIATION TEST

Instructions for Administering

See that each pupil is comfortably seated at a desk or writing table and is provided with a pen or a pencil which is in good working condition. Tell the pupils that you are going to give them an unusual but very interesting "word exercise." Provide each pupil with a copy of the test folder. Direct the pupils to fill in the blanks at the top of the page. Answer any questions that may be asked about what to write after "Grade or Class."

As soon as everyone has filled the blanks at the top of the page, read slowly and distinctly the following explanation:

"This is an exercise in which you are to tell the next word you think of after each word that I pronounce. I shall say a word and show you the same word printed on a card. You are to think at once of some other word. Any word will do, if it is the first word which comes to your mind after I pronounce my word.

"For the first few times, you may look in the printed list for the word that has come to your mind. If your word is in the printed list, make a check mark before it. If your word is not in the printed list, write it on the dotted line at the bottom of the column of printed words. Be sure to check or write the first word you think of after I tell and show my word."

"Now we are ready to begin. I shall tell you and show you my word, and then as soon as some other word comes to your mind, you may write it or check it and hold up your hand. Under 'Word A' in the left-hand column, check or write the first word you think of when I say the word 'CROWD.'" Be sure to pronounce the word distinctly and to show the right card.

"Look down the column under 'Word A' and check the word you thought of, if you can find it. If it is not in the list of printed words, write it on the blank at the bottom of the list, and then hold up your hand to show that you are ready for another word."

TESTS AND MEASUREMENTS

Pause until all but two or three hands are up. Give these pupils extra help and explanations, if necessary. "Do you all understand what to do when I pronounce a word? First, you think of another word, and then you either check or write your word in the proper place and hold up your hand. Are there any questions?" Answer any questions as briefly as possible.

"Under 'Word B' check or write the first word you think of when I say the word 'DIVIDE.'" Show the card with the word on it. "Hold up your hand as soon as you have finished."

"Under 'Word C' check or write the first word you think of when I say the word 'STAR.'" Show the proper card.

"Now turn the page. Perhaps you had better fold the page back so that you can see only the first three columns of lines, numbered from 1 to 75."

"From now on you are to *write* the first word you think of after I give you my word. Be sure to write your word on the dotted line after its proper number. If you don't know just how to spell your word, write it as you think it sounds. This is not a spelling test. Be sure to write the first thing that comes to your mind after you hear my word."

"After the number 1, in the left-hand column, write the first word you think of when I say the word 'LAMP.'" Pronounce each word distinctly and show it on the flash card. Allow about ten seconds, if necessary, before going on with the next word. A careful watching of the slow writers will help determine the proper time. If a following word is pronounced when 90 per cent have finished writing the response to the preceding word, no undue hastening of the slow writers is involved.

For the first eight or ten words, use the formula: "After the number 2, write the first word you think of when I say the word 'RICH.'" Be sure to give the right word and to show the right card for each number.

After the routine has been established, the formula may be shortened to: "For number 10, the word is SCHOOL."

Two or more class periods may be required for completing the two hundred items in the test. Allow at least one rest period during the test.

retained for the three illustrative stimulus words with which the test begins. The first page of the 8½x11-inch test folder is reproduced herewith, together with the top of the second page and the accompanying instructions to the examiner.

The ratings by teachers were obtained in the manner described on page 349, except that each teacher followed a different order in her ratings of the eight traits. This variation in the order of trait ratings was made to break up any constant error due to having all pupils rated first for "Courtesy," second for "Fair Play," etc.

ADMINISTRATION OF TESTS IN 1932

STIMULUS WORDS FOR SOCIAL SCIENCE WORD ASSOCIATION TEST

No.	Word	No.	Word	No.	Word	No.	Word
1	lamp	51	cottage	101	wait	151	white
2	rich	52	eagle	102	wild	152	water
3	ride	53	hurt	103	over	153	silly
4	age	54	kitchen	104	note	154	sad
5	dress	55	pain	105	difficult	155	cold
6	stranger	56	platform	106	desire	156	child
7	sundown	57	light	107	sickness	157	hit
8	night	58	lion	108	shock	158	hate
9	ocean	59	crowd	109	black	159	vacation
10	school	60	dance	110	band	160	try
11	music	61	tire	111	seed	161	butter
12	needle	62	train	112	once	162	bread
13	restless	63	friendly	113	mountain	163	slow
14	river	64	habit	114	milk	164	sky
15	poor	65	smooth	115	evening	165	castle
16	prison	66	soldier	116	eating	166	camp
17	rain	67	king	117	thief	167	mischievous
18	range	68	learn	118	swift	168	love
19	mother	69	flag	119	home	169	sweet
20	name	70	form	120	admire	170	street
21	church	71	join	121	companion	171	cotton
22	citizen	72	joy	122	cave	172	correct
23	pleasure	73	fair	123	pure	173	man
24	prefer	74	fault	124	sorrow	174	make
25	square	75	country	125	sing	175	stem
26	stage	76	deep	126	dark	176	spider
27	loud	77	snake	127	current	177	protect
28	marble	78	star	128	fairy	178	power
29	command	79	queer	129	exercise	179	coat
30	cow	80	college	130	journey	180	choose
31	stove	81	control	131	island	181	must
32	summer	82	fond	132	foot	182	mouth
33	listen	83	honor	133	fire	183	red
34	luck	84	table	134	laugh	184	quiet
35	boy	85	easy	135	sour	185	priest
36	captain	86	enjoy	136	soft	186	pray
37	simple	87	master	137	hair	187	rough
38	sleep	88	moon	138	give	188	ring
39	blue	89	new	139	travel	189	need
40	butterfly	90	rose	140	tobacco	190	nature
41	trip	91	aim	141	date	191	science
42	useful	92	bat	142	cut	192	salt
43	handle	93	sheep	143	long	193	nut
44	hay	94	short	144	like	194	nice
45	cheese	95	delighted	145	playing	195	Sunday
46	city	96	diamond	146	people	196	stomach
47	rush	97	narrow	147	lazy	197	dream
48	silk	98	old	148	important	198	beat
49	watch	99	walk	149	earth	199	dog
50	whistle	100	young	150	doctor	200	dislike

TESTS AND MEASUREMENTS

SECTION VI

CHARACTER STATUS OF THE GROUPS TESTED

If the various measures developed from the responses of pupils in North Carolina and in Washington, D. C., possess any practical significance as indexes of social attitudes or character, it seemed reasonable to suppose that classes elsewhere would differ from each other in their mean scores in accordance with known differences in their social backgrounds and training. An attempt was made, therefore, to locate and test classes of about seventh-grade maturity in which the pupils were believed to possess highly desirable social attitudes and other classes of equivalent maturity in which there was reason to believe that the pupils possessed undesirable social attitudes. Unless the contrast in known attitudes among these classes were reflected by statistically significant differences in their test scores, there could be little hope of attaining ultimate success in developing practical measures of social attitudes or character from responses to a free association test.

One of the outstanding efforts at character development in the elementary schools is that of the Ethical Culture Schools of New York City. Through the kindness of the director, Dr. V. T. Thayer, three sixth-grade classes in this school were given the free association test. The special training which these classes had received has been briefly described for us by Miss Ethel C. Bratton, Principal of the Elementary Department:

"In the Elementary Department, as in all of the departments of the Ethical Culture Schools, the pupils have a definite course of study in moral education. This course is given by means of stories, true and imaginary; by the study of the life and work of fine characters, past and present; and by the discussion of ethical problems which arise in the study of current events as well as in the daily lives of the children. In this way the pupils get a background of knowledge, understanding, and feeling, as well as an attitude of thoughtfulness toward moral questions and behavior.

CHARACTER STATUS OF THE GROUPS TESTED

"There are many opportunities given in the school for the children to practice right ethical principles as well as common courtesies, through their cooperative government, philanthropic work, and freedom in choice and initiative in classroom work."

Among the agencies which have been most active in stimulating attention to character education in the public schools is the National Child Welfare Association. Dr. Frank Astor, of this organization, suggested that Public School No. 93 in New York City would probably show as large effects of character education as any other school available. Through the kindness of various officials of the New York City Public Schools, permission was obtained to examine pupils in a number of schools, including No. 93. Four sixth-grade classes of this school were given the association test. The Knighthood of Youth program, which was being used in P. S. No. 93, may be described briefly as follows:

Knighthood of Youth is a club program for pupils of the elementary grades, emphasizing and building upon the natural interests which children have in stories of knights and ladies in medieval castles. Each school room using this program organizes itself into a club, which adopts its own colors, songs, motto, codes, and the like, and which attempts to build itself a record of socially desirable deeds or accomplishments. Each achievement is graphically recorded on a wall chart as one more stone in the wall of their castle. Socially desirable behavior is held up to the pupil as an ideal of modern knights and as a duty to his club.

Through the cooperation of Mr. K. H. Thompson of Broadwater, Nebraska, we were supplied with the results of the association tests for five sixth-grade classes of the public schools at Scottsbluff, Nebraska. We are informed that the State Department of Public Instruction of Nebraska has been stimulating the use of the Knighthood of Youth program through the publication of books and pamphlets since early in 1930, and that the schools at Scottsbluff have been participating in the program. It should be remembered, however, that the pupils in these classes had not been under the influence of this statewide program for more than a few years at the time the association tests were administered.

TESTS AND MEASUREMENTS

Professor R. W. Fairchild of Northwestern University suggested to us that some unusually effective character training was being given in one of the junior high schools of Kenosha, Wisconsin. Through the kindness of Superintendent G. F. Loomis, we were able to examine 112 seventh-grade pupils in this school. Mr. Loomis describes the school as follows:

"This school, which has been in existence for seven years under the principalship of the same man, had devoted particular interest to the development of school ideals, pupil responsibility, and pupil growth through pupil participation. The principal and teachers have been particularly sympathetic and have felt that they were teaching children rather than subject matter. The set up of the school has provided for a measure of pupil responsibility in matters of direction and discipline, has devoted a portion of each day as an activity period to self-directed and self-controlled activities, clubs, etc. The children come quite largely from foreign and unfavored homes."

Superintendent Theodore Saam of Elgin, Illinois, assisted the authors in securing responses from 311 seventh-grade pupils in his schools. Regarding the special training these pupils had received Superintendent Saam wrote as follows:

"Prior to 1930 there was a very deliberate and definite program of character education in the Elgin schools which was carried on more or less successfully in all the elementary grades. The program provided that the first ten minutes of each day be devoted directly to character education. In the fall of 1930, however, the teachers were not interested enough to continue this type of instruction, although it was approved and permitted. Since 1930 there has been no specific program of character apart from the continuous training of character in tasks assigned and guiding the life of a child during the time he is under the jurisdiction of the teachers."

There appeared in the *New York Times* for Sunday, August 28, 1932, an article regarding the work of the public schools of Norwalk, Conn., under the headline "City Molds Character." In the article Superintendent John Lund wrote as follows:

CHARACTER STATUS OF THE GROUPS TESTED

"Our classroom techniques may be briefly summarized as emphasizing pupil activity, with the attention of the teacher focused on individual development in terms of an integrated personality rather than upon the acquisition of book knowledge and skills."

Mr. Lund very kindly allowed the authors to examine 462 seventh-grade pupils in the Norwalk schools, but in doing so he called attention to two other statements in the *Times* article:

"We have consciously tried to avoid anything that could be construed as a definite character education plan." "We make the emphasis on character development a major preoccupation for all teachers in all parts of the educational scheme."

In order to have a sampling of pupils who were known to be deficient in desirable social attitudes, permission was obtained to examine pupils in certain parental and probationary schools in New York City. The boys in these schools had all been committed to them by court order because of misdemeanors or delinquency of one sort or another. The sixth, seventh, and eighth grades of Public School No. 111, a parental school in Flushing, of Public School No. 61, a probationary school in Brooklyn, and of Public Schools Nos. 37 and 174, probationary schools in Manhattan, were therefore examined. These parental and probationary schools were assumed to contain children of less than average social character, while Public School No. 93 and the Ethical Culture School were assumed to contain children of more than average social character. At the time the tests were administered the authors were uncertain how they should classify the character status of pupils tested in schools outside of New York City.

SECTION VII

THE SCORING OF THE TESTS USED IN THE SURVEY PROGRAM

For the free association test to become a generally serviceable instrument, it is necessary to reduce the labor cost of scoring it. Roughly, a sample of seven hundred seventh-grade pupils will

TESTS AND MEASUREMENTS

give from fifty to one hundred and fifty different responses to a single stimulus word. A larger sample would yield additional responses. Each different response may be expected to hold a unique story as to the mean outlook and associative tendencies of those so responding. If to the stimulus "pirate" one child responds "knife" and another "past," it may be assumed that their interests and attitudes differ, or rather that the mean tendencies of the members of an entire group, all of whom respond "knife," differ from those of a group all of whom respond "past." Back of each response word is a correlative associative process characterizing the individual making the response. If we can but discover this mediating view, and credit or debit the response word in harmony therewith, we have to this extent a means of scoring a new subject who upon being given the stimulus word makes the same response. To grade two hundred response words for each of two hundred stimulus words upon each of eight traits calls for the determination of some $(8 \times 200 \times 200) = 320,000$ debits and credits, each one a statistical valuation based upon the trait ratings of the subjects giving the response in question in comparison with subjects in general. This is not only an enormous task to accomplish, but, once done, it yields so cumbersome a scoring scheme as to militate against its use. To abridge the process, a grouping of responses (later found to be too radical) was resorted to:

- A. The most frequent responses to a single stimulus, but never more than forty-four in number, were evaluated separately, yielding eight different scores for the eight traits studied, and
- B. Other responses, infrequent, but in the judgment of the chief clerk in charge judged to be very similar, were grouped together. For example, if to the stimulus "pirate" ten out of the seven hundred responded "weapon" and one or two responded "knife," these would have been classified together.
- C. Failures to respond were treated as constituting a single class.
- D. All other responses were treated as a single class, designated "unique."

THE SCORING OF THE TESTS

This last procedure (D) was dictated by considerations of economy and time, and proved unfortunate. It was, of course, even at the start thought that this procedure would tend to lessen the reliability of the instrument, but it was only after most of the statistical work had been done that it was discovered that it led to certain systematic errors as well.

The following is offered as a surmise as to the nature of this error: When the members of the standardization group, composed of typical seventh-grade school children, gave a unique response, it was ordinarily because of a high type of association, *e. g.* "pirate" and response "racketeer," or again, "pirate" and response "law-abiding," with the result that the mean trait ratings were high, thus leading to the final result that a unique response was generally given a positive credit for a majority, or in some cases for all of the eight traits. Thus, on the whole, unique responses were a good sign in the case of the standardization group. However, when delinquent boys in probationary and parental schools gave unique responses, they were probably consequent to a different type of mental process. Many of these responses were not only unique in the sense that other children (of the seven hundred studied) did not give them, but in the sense that they were utterly unrelated as judged by an intelligent adult, *e. g.* "pirate" and the response "street," or "pirate" and response "go." (One delinquent boy for a stretch of some fifty words responded in order with "I," "don't," "know," "I," "don't," "know," "I," "don't," "know," etc.)

To give a unique response made by a delinquent boy a generally high trait credit does not seem to be sound, though to give a unique response such a credit when the subject is a typical school child may be sound. At the time this hiatus in the significance of unique responses was discovered, it was far too late to attempt to remedy it, except as a partial remedy might be accomplished by rescoring all papers omitting all unique responses. (Even this latter task proved to be an undertaking of several months' time.) Obviously the real remedy is to be found in the elimination or the lessening of the unique-response class by (1) determining scores

TESTS AND MEASUREMENTS

for many words now classified as unique, and (2) grouping the rest in a number of carefully defined classes rather than bunching them together as of a single type. One such sub-class may well be composed of rhyming and other responses based upon the sound of the stimulus word. For example, "pirate" and the response "pie," or "pirate" and the response "irate."

The first of the attached lists, in Table V, of unique responses contains samples drawn from pupils of the Ethical Culture School in New York City, who may be considered above average in such character traits as are here studied. Five response words were taken from the responses of each of twenty-five pupils taken at random. The second list, gotten in a similar manner, contains responses from pupils of probationary school No. 37 of New York City. An examination of single stimulus-response connections will scarcely reveal significantly different group tendencies, but remembering that the total trait score of a pupil is a summation of 198 credits or debits, it will be seen that a very small systematic difference will sum up to a substantial total. Admitting that the degrees of excellence of most of the individual associations are indistinguishable from group to group, still the writers judge that there is a preponderance of high type associations in the Ethical Culture group and of low type associations in the probationary school group. A study of this matter may be aided by an examination of the differences in unique responses from the two groups to the same stimulus words. Table VI is a random sample of such responses.

TABLE V
UNIQUE RESPONSES, ETHICAL CULTURE SCHOOL

Stimulus	Response	Stimulus	Response
age	how	smooth	silverware
music	rhythm	star	shawl
name	everyone	college	team
marble	column	master	sir
luck	skill	new	address
blue	paste	young	singing
whistle	pipe	mountain	sand
kitchen	kitchenette	thief	liar

THE SCORING OF THE TESTS

TABLE V (Cont.)

UNIQUE RESPONSES, ETHICAL CULTURE SCHOOL

Stimulus	Response	Stimulus	Response
cave	drown	form	golf
pure	puries	college	where I'm going
current	seed	honor	my
fire	glory	sickness	lifetime
important	laws	shock	burst
silly	soossy	once	an
cold	coat	cave	room
child	chide	fairy	light
sky	sly	exercise	reduce
castle	crag	hair	barber
love	mush	important	mail
sweet	strawberry	camp	out
street	policeman	protect	antagonize
must	really	rough	gentle
red	bull	ring	bracelet
priest	Marquette	need	useless
science	Miss Klaer	nice	disagreeable
church	cathedral	age	children
citizen	French Revolution	school	friends
marble	immy	river	floods
sleep	done	rain	spring
join	help	trip	leave
kitchen	place	whistle	reed
flag	Betsy Ross	cottage	lovely
queer	quirios	eagle	nation
control	levee	kitchen	dinette
new	me	platform	dias
bat	haunted	smooth	iceing
walk	brisk	form	art
note	secret	fair	goodsport
pure	sweet	college	women
fairy	glade	control	laughter
island	raft	new	present
important	main	companion	people
earth	cool	current	sea
quiet	tranquil	fire	flower
rough	splinter	sour	food
nut	tap	lazy	unuseful
nice	mice	important	quick
restless	uneasy	camp	groupe
hay	loft	choose	among
city	busy	mouth	speak
whistle	mouthpiece	ride	with
eagle	crag	stranger	relative
train	smash	sundown	bar
habit	have it	needle	spin
king	me	restless	calm

TESTS AND MEASUREMENTS

TABLE V (Cont.)

UNIQUE RESPONSES, ETHICAL CULTURE SCHOOL

Stimulus	Response	Stimulus	Response
citizen	home	diamond	glisten
command	hard	wait	leave
summer	joy	wild	meek
luck	inherit	sickness	astonish
simple	nice	eating	dineing
butterfly	dainty	home	comfortable
kitchen	smoke	note	keynote
light	gleam	admire	ability
lion	viscious	sing	softly
dance	group	exercise	scales
train	learn	foot	weary
habit	rotien	long	teditious
form	games	lazy	daisy
new	brand	earth	God
aim	standard	water	fall
young	pep	cold	quits
wild	fierce	hate	bitter
sickness	not well	bread	worms
pure	real	love	pansy
hair	thread	man	many
important	needed	make	finnist
child	human	stem	tide
hit	collide	must	mess
slow	speed	ring	rasey
coat	aparall	nut	cake
choose	team	Sunday	7 day
must	you	dog	leash
red	corpasciles	stranger	fromafar
quiet	evening	poor	homeless
priest	cardinals	rain	shadows
rough	noise	name	everyone has
nature	dispossession	citizen	some cannot be
poor	less fortunate	marble	color
salt	spice	luck	much
citizen	belongs to city	useful	hand
stove	furnace	hay	slide
dance	performance	city	old
smooth	glide	crowd	heat
form	method	form	a
fair	rain	join	Red Cross
country	far	country	Scarsdale
deep	earth	honor	your
college	coed	master	overseer
table	floor	wild	tangled
easy	very	band	strip
enjoy	myself	pure	cool
moon	change	fairy	light

THE SCORING OF THE TESTS

TABLE V (Cont.)

UNIQUE RESPONSES, ETHICAL CULTURE SCHOOL

Stimulus	Response	Stimulus	Response
exercise	lesson	summer	pretty
foot	bottom	uck	nothing
give	part with	useful	airplane
important	needful	rush	subway
sky	horizon	whistle	grass
castle	fortress	platforms	convention
camp	Montana	habit	want
man	being	form	Fieldstar
mouth	cheek	country	sweet
red	pink	snake	big
need	to	star	lone
science	astronomy	control	wheel
stranger	shepherd	honor	beat
ocean	cold	aim	fine
music	flagillette	delighted	apisative
needle	blood	note	that
restless	uncomfortable	shock	prove
river	sailing	black	mist
poor	no nothing	band	theator
rain	terreble	seed	blossom
range	what	once	life
mother	happy	swift	eagle
square	place	companion	brother
loud	ugly		

UNIQUE RESPONSES, PUBLIC SCHOOL 37

stranger	hate	island	prison
name	right	shock	burn
pleasure	ease	laugh	straight
boy	boat	travel	tony
city	open	important	sereas
luck	prosper	mischief	done
blue	schady	correct	dome
trip	picnic	power	pull
hurt	bandage	citizen	countrymen
friendly	acquainted	prefer	go
queer	clear	square	even on four sides
college	trade	summer	fun
control	people	simple	nonothing
easy	swim	blue	cloud
enjoy	him	habit	stop
over	jay	queer	nuts
desire	insist	wild	manica
sickness	reurt	once	quarte
fairy	straight	admire	dogs

TESTS AND MEASUREMENTS

TABLE V (Cont.)

UNIQUE RESPONSES, PUBLIC SCHOOL 37

Stimulus	Response	Stimulus	Response
current	nice	doctor	seuence
exercise	bones	sad	cannot
laugh	mouth	cold	very
soft	cat	try	what
long	pipe	sky	eve
people	eat	sweet	potato
sing	professor	make	cloth
foot	amutate	protect	armed
cut	illegible	nice	pal
people	bing	dog	fake
lazy	nogood	music	soft
white	plain	poor	dressed
hate	ate	name	somebody
try	rye	blue	sticky
butter	rye	habit	to keep on
slow	rotten	smooth	cleab
mischiev	chief	joy	joy
street	earth	form	formation
correct	correction	country	fresh air
stem	unsterb	star	tree
power	unpower	honor	him
mouth	mother	easy	shoe
rough	unrough	moon	half moon
beat	fast	rose	agril
ride	move	black	polish
music	entertainment	band	misc
restless	movies	pure	reald
name	who	sing	micedar
cow	cud	dark	dick
boy	boy	journey	visitor
train	ecing	island	sourrand
habit	some	foot	far
form	do	soft	mice
fault	play	people	sombod
country	go	cold	frost
fond	save	hate	sea
moon	look	vacation	abest
wild	bite	cotton	goings
difficult	do	stem	end
shock	fish	need	bring
sorrow	surou	salt	soft
current	courage	nut	dope
exercise	tricks	nice	darling
sour	sour	stomach	guts
soft	moos	dream	magic
long	story	beat	leg
like	others	dislike	nogood

THE SCORING OF THE TESTS

TABLE V (Cont.)

UNIQUE RESPONSES, PUBLIC SCHOOL 37

Stimulus	Response	Stimulus	Response
stranger	conral	beat	tree
stage	plarforel	music	radio
luck	close	restless	calm
useful	allwdes	name	unknown
whistle	signal	prefer	prefect
form	march	stage	baconly
country	live	stove	radiator
honor	complement	boy	son
easy	to do	handle	earn
enjoy	time	hurt	curt
master	dracula	light	bad
diamond	cool	train	subway
wait	halt	habit	think
wild	lion	join	going
once	nremenber	fair	chance
fairy	sissy	fault	so
exercise	stretch	star	arice
people	citizen	queer	to give in
lazy	him	honor	privilege
sad	vey losjsmen	table	place to lean on
child	daugther	moon	part of the earth
sweet	soft	aim	to get
power	over you	sheep	fleece
need	circle	diamond	coal
beat	rub	seed	orange
night	night	mountain	fell
church	private	home	lodging
prefer	among	sing	solo
square	equal	give	something
light	bad	like	long
learn	play	lazy	lazey
country	children	slow	sour
honor	to be	street	tar
easy	thing	sorry	science
new	just bought	pray	praise
young	future	rough	tumble
shock	noise	dream	feeling
pure	sterilized	beat	hot
cut	split	sundown	late noon
water	refreshing	command	charge
hate	unlike	city	down
slow	unfast	rush	clean
sky	work	watch	lookout
stem	edge	friendly	partner
quiet	quiteful	smooth	parm
nature	live	form	change
dream	dream	fair	even

TESTS AND MEASUREMENTS

TABLE V (Cont.)
UNIQUE RESPONSES, PUBLIC SCHOOL 37

Stimulus	Response	Stimulus	Response
snake	nawwow	cotton	salt
short	not much	protect	body
narrow	skinny	must	due
old	long lived	beat	animal
difficult	unpr	musical	musical
once	hurry	make	make
sour	not good	summer	enjoyment
tobacco	war	science	still
long	give	ride	fruit
like	elect	mother	us
lazy	weak	stove	rock
hit	punch	listen	study

TABLE VI
RANDOM UNIQUE RESPONSES FROM ETHICAL CULTURE AND
DELINQUENT GROUPS

Stimulus	Ethical Culture	Probationary
ride	with	move
stranger	relative	hate
"	shepaerd	conral
"	fromafar	
sundown	bar	late noon
music	rythym	entertainment
"	flagilette	radio
"		soft
restless	uneasy	movies
"	uncomfortable	calm
"	calm	
poor	no nothing	dressed
"	less fortunate	
"	homeless	
name	everyone	unknown
"	everyone has	right
"		somebody
church	cathedral	private
citizen	home	countrymen
"	French Revolution	
"	belongs to city	
command	hard	charge
stove	furnace	rock
"		radiator
summer	joy	fun
"	pretty	enjoyment

THE SCORING OF THE TESTS

TABLE VI (Cont)

Stimulus	Ethical Culture	Probationary
luck	nothing	prosper
"	skill	close
"	much	
"	inherit	
simple	nice	nothing
blue	paste	cloud
"		schady
trip	leave	picnic
useful	hand	allwdes
city	busy	open
"	old	down
rush	subway	clean
whistle	mouthpiece	pipe
"	signal	
"	reed	
light	gleam	bad
train	smash	ecing
"	learn	subway
habit	have it	stop
"	rotten	think
"	want	some
"		to keep on
smooth	silverware	cleab
"	glide	parm
"	iceing	
form	golf	do
"	art	formation
"	Fieldstar	march
"	games	change
"	a	
"	method	
join	Red Cross	going
"	help	
fair	goodsport	chance
"	rain	even
country	far	go
"	Scarsdale	children
"	sweet	fresh air
"		live
snake	big	nawwow
star	shalw	tree
"	lone	arice
queer	quirios	to give in
"		clear
"		nuts
college	coed	trade
"	team	
"	women	
control	wheel	people

TESTS AND MEASUREMENTS

TABLE VI (Cont.)

Stimulus	Ethical Culture	Probationary
control	levee	
"	laughter	
honor	my	him
"	your	complement
"	beat	privilege
"		to be
table	floor	place to lean on
easy	very	to do
"		swim
"		shoe
enjoy	myself	him
master	overseer	dracula
"	sir	
moon	change	part of the earth
"		look
"		half moon
new	present	just bought
"	me	
"	brand	
aim	standard	to get
diamond	glisten	cool
young	singing	future
wait	leave	half
wild	fierce	manica
"	meek	bite
"	tangled	lion
sickness	astonish	reurt
"	life time	
"	not well	
shock	burst	burn
"	prove	noise
"		fish
black	mist	polish
band	theator	miscl
"	strip	
once	an	quarte
"	life	hurry
"		nremenber
mountain	sand	fell
pure	puries	reald
"	real	sterilized
"	sweet	
current	seed	nice
"	sea	courage
fairy	glade	straight
"	light	sissy
exercise	reduce	bones
"	scales	tricks
"	lesson	stretch

THE SCORING OF THE TESTS

TABLE VI (Cont.)

Stimulus	Ethical Culture	Probationary
island	sourrand	prison
"	raft	
foot	bottom	amurate
"		far
sour	food	sour
long	teditious	give
"		pipe
lazy	unuseful	no good
"	daisy	lazey
"		him
important	needful	sereas
"	laws	
"	quick	
"	needed	
water	fall	refreshing
cold	quite	very
"	coat	
child	human	daughter
"	chide	
hate	bitter	unlike
"		ate
"		sea
slow	speed	rotten
"		sour
"		unfast
sky	sly	eye
"	horizon	work
sweet	strawberry	potato
street	policeman	earth
make	finnist	make
"		cloth
stem	tide	edge
"		unsterb
"		end
protect	antagonzise	armed
must	mess	due
"	really	
"	you	
mouth	speak	mother
quiet	tranquil	quiteful
rough	splinter	unrough
"	noise	tumble
"	gentle	
need	useless	bring
science	Miss Klaer	still
nice	disagreeable	darling
"	mice	pal
dog	leash	fake

TESTS AND MEASUREMENTS

Though the processes of association are so subtle that it is not claimed the reader or the authors should be able to do so, nevertheless it may be that one could judge by perusing these lists whether the character traits indicated by the one set of responses are generally higher than those indicated by the other.

The authors believe that the preceding lists give support to the explanation offered, but this explanation is distinctly a surmise. It would seem that a grading of the unique responses of the Ethical Culture group or of typical school groups according to the best scoring scheme now available, wherein unique responses tend to receive positive credit, would be sound, and that omitting all grading of unique responses for the probationary groups would not be unfair to them. This dual procedure would lead to much more substantial group differences than those reported in Table X. The writers have not followed such a dual standard of grading, but they believe that the differences shown in Table X are decided understatements of the differences which would have been revealed had a more accurate means of grading the unique responses been at hand.

Another, though less important, shortcoming in the test as given and as credited lies in the credits and debits given for "omitted" responses. No clearly systematic difference between delinquent and normal groups upon this basis has been discovered, except that more such omissions were found in the case of the delinquent groups. This undoubtedly makes for lessened reliability, though perhaps it does not lead to a systematic error. This defect can be remedied by greater care in test administration. A teacher can observe that a child is writing nothing, but he is unlikely to discover while the test is in process that he is writing "I," "don't," "know," etc.

Improvements both in the administration of the test and in its scoring, which will make for greater reliability and validity, seem to be definitely and specifically indicated as a result of the work to date.

THE SCORING OF THE TESTS

THE SCORING TECHNIQUE FOLLOWED (HOLLERITH)

The task of assigning credits to different responses to each of the 198 stimulus words constituting the final test was a long undertaking. It finally resulted in tables, one for each stimulus word, giving for each of the more frequent responses eight different scores, one for each trait. Certain less frequent responses, judged to be very similar in the idea involved to certain of the more frequent responses, were given the same credits as those attached to the more frequent words. In the form in which they were first used in the actual scoring of pupils, these tables were typed on large cardboards, fifty to a board, so that the scorer was called upon to handle four of them in the scoring of a single test. For each response word the scorer would first find the response and then write down manually in appropriate columns the eight different trait scores, except that zero scores were omitted. These columns were then added, yielding eight trait scores for the pupil. The time to accomplish this with expert scorers was about one hour per test of 198 words.

When the subsequent testing under the auspices of the Carnegie Foundation for the Advancement of Teaching was undertaken, a serious effort was made to reduce the cost of scoring, and in this the authors had the generous cooperation of the International

RESPONSE		CHARACTER RATING										STIMULUS	
		1	2	3	4	5	6	7	8	9	10		
A-RESPONSE	NO	0	0	0	0	0	0	0	0	0	0	STIMULUS	NO
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
B-RESPONSE	NO	0	0	0	0	0	0	0	0	0	0	STIMULUS	NO
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		
	NO	0	0	0	0	0	0	0	0	0	0		

TESTS AND MEASUREMENTS

Business Machines Corporation, who adapted one of their tabulators to the problem in hand and built a special reading unit. The eight scores attached to a single response were punched upon one half of the card shown on previous page.

The other half of the card was used for the next word in the alphabetical list of responses to the same stimulus word. The scores were multiplied by an appropriate factor to reduce their variability, enabling representation upon a 9-unit scale: 4, 3, 2, 1, 0, —1, —2, —3, —4. For each of the eight traits these reduced scores were punched upon the response score card, one column per trait being required.

For each stimulus word there were not over twenty-two scoring cards, yielding not over forty-four different sets of gradings, as each end of the card gives a separate grading. For each stimulus word, these response score cards were filed in a separate holder, so that the ends bearing the response words and arranged alphabetically were staggered and at all times visible.

In scoring the response of a pupil, the scorer would note the pupil's written response; draw the corresponding response card; enter it into the automatic reading device, one end if for the first word and the other end if for the next word as typed on the top of each card; press a lever, which would result in eight addends being printed and entered into the appropriate tabulators; remove the card from the reading device, and reinsert it in its original position in the file of response words. This was then repeated for the second word, etc., through the list of 198. Pressing a total button then gave the printed sums of the eight columns.

In the case of expert machine operators, some 500 responses, or $2\frac{1}{2}$ papers, were scored per hour. Though the time required to score a test is but two-fifths the former time, it is still a considerable undertaking. With the increase in the number of response words scored, which is an important improvement definitely called for, the time cost of scoring a paper would probably be about thirty minutes rather than, as now, about twenty-four minutes. With this type of test, the scoring cost is, and is likely to remain, the major expense, but such expense as has been indicated should

GROUP DIFFERENCES

scarcely be considered prohibitive if fairly trustworthy measures of character result therefrom.

SECTION VIII

GROUP DIFFERENCES

A number of points connected with the scoring of the free association test and with its reliability should be made clear before a discussion is entered into of differences in performance of classes from probationary and parental schools on the one hand, and of classes from schools having had some sort of character training on the other. The accompanying Table VII gives the mean test scores for 3 sixth grade sections of the Ethical Culture School of New York, and for 7 sixth to eighth grade sections from parental and probationary schools in New York.

TABLE VII
MEAN SCORES

School	Class	No. of	Cour-	Fair	Hon-	Loy-	Mas-	Poise	Prop.	School
		cases	tesy	Play	esty	alty	tery		Rts.	Drive
Ethical Culture	6 ¹	20	14.8	15.7	21.6	15.2	4.5	14.4	19.8	18.6
	6 ²	24	17.8	20.0	24.6	19.8	4.8	19.0	23.1	21.9
	6 ³	21	15.5	21.0	24.2	14.6	5.6	20.0	19.3	17.0
P. S. 111 Parental	6	50	3.8	8.7	8.0	9.3	4.8	8.3	8.3	.7
	7	60	3.6	8.9	9.5	8.7	8.7	7.3	7.0	4.9
	8	46	3.7	10.1	11.6	8.3	5.2	5.0	7.4	8.2
P. S. 174 Probationary	6	19	.7	-1.8	.3	1.0	4.8	-.2	-.9	-3.4
	7	12	3.3	2.5	2.1	.3	4.1	-2.3	3.2	1.8
	8	17	2.2	2.8	3.8	7.6	8.2	.9	5.4	2.0
P. S. 37 Probationary	6	48	-3.1	.2	-5.8	3.0	5.7	-5.3	-1.8	-12.3

The standard deviations of the distributions for the respective classes have not been calculated, but are approximately as follows, which are the average values for the classes of Table X:

TESTS AND MEASUREMENTS

TABLE VIII
APPROXIMATE STANDARD DEVIATIONS OF CLASSES

Courtesy	Fair Play	Honesty	Loyalty	Mastery	Poise	Prop. Rts.	School Drive
10.21	10.24	11.64	8.92	8.08	11.52	10.82	11.05

Table VII may be summarized into the table herewith.

TABLE IX
MEANS AND APPROXIMATE STANDARD DEVIATIONS OF
MEANS FOR ETHICAL CULTURE AND PROBATIONARY GROUPS

	Size of Sample	Courtesy	Fair Play	Honesty	Loyalty	Mastery	Poise	Prop. Rts.	School Drive
Ethical Culture	65	16.1	19.0	23.5	16.7	5.0	17.9	20.9	19.3
Estimated Standard Error of Means		1.3	1.3	1.4	1.1	1.0	1.4	1.3	1.4
Probationary groups	252	2.1	6.7	5.2	6.6	6.2	3.2	4.8	.4
Estimated Standard Error of Means		.6	.6	.7	.6	.5	.7	.7	.7

It would seem that the superiority of the Ethical Culture group for every trait except that of mastery was indubitably established. It, however, must be pointed out that there is a vitiating factor revealed by these means which had not been anticipated, so that differences between the groups are not as well established as this table would indicate. Otherwise expressed, the differences between the groups are not to be attributed altogether to differences in the traits as labelled, but in part to differences in the effective administration of the test. It is to be noted that the mean scores for the delinquent groups are positive. This is contrary to expectation because the mean score in the case of the normal children involved in the construction of the scoring scheme is zero, and it is hardly to be expected that these delinquent boys are superior in these eight character traits to the children of the standardization group. After considerable investi-

GROUP DIFFERENCES

gation, the reason for this was established, as mentioned in Section VII, as due to the positive credit received for unique responses, and to a very small degree to omitted responses. Accordingly all papers were rescored, omitting all pupils having 40 or more omitted responses, and omitting the scores attached to all unique responses. It may be believed that the more nearly normal, that is "like the subjects of the standardization group," the more unfair to a child is this procedure. That is, the procedure of omitting unique responses is probably more unfair to the Ethical Culture group than it is to the delinquent group. Accordingly, as noted in connection with Table VI, differences in the accompanying Table X between the groups are not only understatements of trait differences, but understatements even of the measures of them that would be obtained were the scoring scheme extended to yield differential scores for responses now classed as unique.

While examining the mean trait scores of Table X, the reader should remember that unique responses have not been scored, with the result that a slightly negative trend is to be expected. The trend would also be negative in the standardization group if scored in this manner.

Let us study this table both by rows and by columns. The children of the Ethical Culture School generally stand high in all traits except Mastery. The girls of P. S. 93 are high in all traits except Mastery, but especially high in Courtesy, Honesty, and School Drive. This may be consequent to the Knighthood of Youth character training which has been in thorough operation at that school, or it may represent a sex difference, for this is the only school measured which was composed entirely of girls. The children of Elgin are quite normal in most traits, but they have slightly low scores in Fair Play and Poise. Those of Kenosha are somewhat high in all traits, including Mastery. Those of Norwalk would seem to be about normal, as judged by the standardization group, though perhaps a little high in Loyalty and School Drive, and low in Poise. Those of Scottsbluff would seem to be normal as judged by the standardization group. Those in the probationary schools

TABLE X
MEANS AND STANDARD DEVIATIONS (ESTIMATED FROM AVERAGE DEVIATIONS)
OF THE SEPARATE CLASSES STUDIED

City	School	Class	Size of Sample	Mean Scores							
				Courtesy	Fair Play	Honesty	Loyalty	Mastery	Poise	Prop. Rights	School Drive
New York	Ethical Culture	6 ¹	20	2.8	-2.4	3.4	2.2	-.6	-2.6	2.5	7.3
		6 ²	24	6.9	3.4	8.0	8.1	.2	3.7	7.3	12.5
		6 ³	21	7.0	8.0	11.1	5.5	2.0	8.5	7.0	9.7
	Average for school		65	5.6	3.1	7.6	5.5	.6	3.3	5.7	10.0
Flushing, N. Y.	Public School 93	6A ¹	41	10.7	2.7	12.7	9.5	.8	2.8	8.2	14.8
		6A ²	34	6.4	4.3	6.9	3.9	2.8	4.4	7.4	8.8
		6B ¹	48	12.2	4.1	9.7	9.5	-.2	5.4	9.7	13.4
	Average for school	6B ²	48	8.3	3.0	7.8	4.0	-.6	2.8	6.3	9.3
New York	P. S. 111 Parental		171	9.6	3.5	9.4	6.8	.5	3.8	7.9	11.6
		6	45	-2.7	-1.8	-1.1	1.1	.9	.3	-1.8	-2.3
		7	55	-3.3	-1.0	.4	.9	6.2	-1.4	-2.6	.9
	Average for school	8	44	-3.7	-1.7	1.1	.2	2.5	-4.9	-3.5	2.3
New York	P. S. 174 Probationary		144	-3.2	-1.4	.1	.8	3.4	-1.9	-2.6	.4
		6	16	-3.0	-8.1	-5.1	-4.8	2.9	-5.4	-6.1	-4.2
		7	10	1.1	-1.6	1.8	-1.8	3.7	-3.3	-1.4	4.7
	Average for school	8	14	-1.6	-2.3	.1	3.2	7.1	-3.5	-.8	-.6
New York	Average for school		40	-1.5	-4.4	-1.6	-1.2	4.6	-4.2	-3.0	-.7

TABLE X (Continued)

City	School	Class	Size of Sample	Mean Scores							
				Courtesy	Fair Play	Honesty	Loyalty	Mastery	Poise	Prop. School Rights Drive	
Brooklyn	P. S. 61 Probationary	6	22	-3.7	-4.8	-2.0	-2.3	.6	-6.5	-2.3	-1.8
		7	34	-.6	-1.3	.6	1.6	1.8	-4.5	-1.0	1.5
		8	30	-9.2	-9.8	-10.7	-3.9	6.6	-13.3	-12.3	-9.0
	Average for school		86	-4.4	-5.2	-4.0	-1.3	3.2	-8.1	-5.3	-3.0
New York	P. S. 37 Probationary	6	30	-5.8	-3.2	-5.2	-2.3	2.7	-5.2	-5.7	-6.9
		7	57	-.1	-2.9	-.3	.6	4.3	-1.8	.1	1.6
		8	61	-2.8	-7.0	-3.7	-.7	5.5	-6.5	-6.0	-2.6
	Average for school		148	-2.4	-4.7	-2.7	-.5	4.5	-4.4	-3.6	-1.8
Elgin, Ill.	Columbia	7A,B Average	27	2.4	-3.2	1.9	.8	.5	-4.3	.9	2.8
			Deviation	8.1	7.0	7.2	5.7	8.1	8.7	7.0	7.8
			Est. σ	10.1	8.7	9.0	7.2	10.2	10.9	8.8	9.9
				Est. σ_M	2.0	1.7	1.8	1.4	2.0	2.1	1.7
Lord	7A,B	19	5.2	-2.2	3.7	3.5	-.7	-4.2	1.4	5.9	
			Est. σ_M	2.1	1.4	2.2	2.4	2.3	2.5	2.0	2.8
	Grant	7A,B	39	-2.0	-5.0	1.5	.3	2.0	-4.9	-4.3	1.9
Est. σ_M			1.3	1.4	1.6	1.4	1.4	1.6	1.5	1.5	

TESTS AND MEASUREMENTS

TABLE X—(Continued)

City	School	Class	Size of Sample	Mean Scores							
				Courtesy	Fair Play	Honesty	Loyalty	Maturity	Poise	Prop. Rights	School Drive
Elgin, Ill.	McKinley	7A,B	34 Est. σ_M	5.3 1.6	-6 1.9	5.6 2.0	4.1 1.9	5 1.3	-1 1.9	2.4 1.9	5.2 1.7
-	Lincoln	7A,B	24 Est. σ_M	1.6 1.6	-2.0 2.1	4.2 1.9	4.0 2.0	1.0 1.5	-2.8 2.1	-9 1.9	5.8 2.2
	Lowne	7A,B	36 Est. σ_M	3.0 2.0	-5.4 1.8	6 2.5	-2 1.7	-1.4 1.5	-8.6 2.0	-9 2.2	5.3 1.9
	Franklin	7A,B	26 Est. σ_M	2.6 2.4	-2.8 2.3	6.1 2.3	3.1 1.9	6 1.5	-5.1 2.5	-6 2.2	5.6 2.3
	Wing	7B	14 Est. σ_M	2.1 2.8	-3.5 2.5	1.2 2.7	0 2.3	-5.1 2.2	-6.0 2.0	-1.2 2.2	3.6 2.6
	Garfield	7A,B	38 Est. σ_M	-2.4 1.8	-3.6 1.6	-1 1.7	3.5 1.7	3.7 1.5	-5.2 1.8	-4 1.9	3.4 1.8
	Sheridan	7A,B	24 Est. σ_M	4.0 2.2	-7.9 2.1	1 3.0	-0.4 2.3	2 1.9	-7.8 2.5	-3 2.0	5.8 2.2
	Washington	7A	20 Est. σ_M	2.4 2.2	-1.6 2.1	2.8 1.9	1.1 2.1	2.0 1.8	-2.6 3.0	6 2.1	.7 1.9
	Washington	7B	10 Est. σ_M	-3.9 3.8	-2.3 3.5	-2.6 4.8	2.2 3.7	2.0 3.5	-2.9 4.9	-6.7 4.2	-1.8 3.2
	Average Mean for Elgin		311	1.7	-3.5	2.2	1.9	7	-4.7	-7	4.0

TABLE X--(Continued)

City	School	Class	Size of Sample	Mean Scores							
				Cour- tesy	Fair Play	Hon- esty	Loy- alty	Mas- tery	Poise	Prop. Rts.	School Drive
Kenosha, Wis.	Central Junior	7A	30 Est. σ_M	4.4 2.1	5.6 1.8	5.3 2.2	5.2 1.2	2.6 1.3	1.4 1.6	3.2 2.0	8.7 2.6
		7A	28 Est. σ_M	2.4 1.6	5.0 1.4	9.4 2.2	5.4 1.5	3.5 1.1	5.2 1.5	2.1 1.7	6.5 2.1
		7B	54 Est. σ_M	4.2 1.1	3.1 1.3	6.2 1.5	6.6 1.1	1.4 1.5	2.9 1.4	2.8 1.5	7.3 1.1
	Average Mean for Kenosha			112	3.8	4.2	6.7	5.9	2.2	3.3	2.7
Norwalk, Conn.	Center	7A1	34 Est. σ_M	6.5 1.6	1.9 2.0	8.8 1.9	6.4 1.5	-4.8 1.4	2.1 1.8	6.3 2.1	11.5 2.2
		7A2	31 Est. σ_M	-.8 2.2	-4.9 1.8	4.9 2.0	2.1 1.7	-3.3 1.7	-3.4 2.2	.8 1.9	8.4 2.0
		7A3	27 Est. σ_M	-7.1 1.8	-7.3 2.0	-8.7 1.9	-1.6 1.4	.7 1.7	-9.6 2.5	-8.9 2.7	-4.6 2.8
	7A4	15 Est. σ_M	-2.0 2.8	-3.5 1.8	-6.9 2.6	.5 2.2	-3.2 1.8	-10.0 2.5	-7.3 2.9	-6.1 3.4	
7B1	26 Est. σ_M	-3.6 2.1	-3.2 2.0	-1.4 2.3	.2 2.7	1.6 1.9	-1.3 2.0	-1.1 2.2	5.6 1.8		
	7B2	14 Est. σ_M	-4.9 2.7	-7.0 3.8	-4.7 3.7	-.5 2.2	-7.1 2.2	-8.6 3.8	-9.1 3.0	-5.2 2.7	
	Average M for Center School			147	-1.3	-3.5	.0	1.7	-2.4	-4.0	-1.8

TESTS AND MEASUREMENTS

TABLE X—(Continued)

City	School	Class	Size of Sample	Mean Scores							
				Courtesy	Fair Play	Honesty	Loyalty	Mastery	Poise	Prop. Rts.	School Drive
Norwalk, Conn.	Roger Ludlow	7A1	23 Est. σ_M	1.0 1.9	-3.4 1.8	4.0 2.1	4.4 1.3	.6 2.4	-3.4 2.4	2.5 1.8	7.1 2.0
		7A2	19 Est. σ_M	-3.7 2.5	-7.2 2.5	-2.9 2.9	-1.5 2.1	.4 2.9	-11.3 2.5	-5.6 3.1	-1.3 2.8
		7B1	23 Est. σ_M	.3 2.1	-4 2.0	-2.6 2.1	-.9 2.4	1.1 1.6	-4.0 2.3	2.5 2.0	6.4 2.0
		7B2	20 Est. σ_M	-2.0 4.3	-4.3 2.6	-6.8 2.8	1.4 2.0	3.0 1.9	-6.6 2.4	-6.0 2.3	1.0 2.6
	Average M for Ludlow School		85	-.9	-3.7	-.4	.9	1.3	-6.1	-1.3	3.6
	Franklin	7A1	40 Est. σ_M	5.6 1.3	1.5 1.5	5.8 1.8	5.0 1.6	-.5 1.3	-.6 2.3	2.1 1.6	9.9 1.4
		7A2	36 Est. σ_M	.6 1.7	-4.5 1.6	-.6 2.0	.3 1.7	2.2 1.1	-.8 2.3	-2.3 1.7	3.4 1.9
		7A3	34 Est. σ_M	-2.8 1.7	-5.4 1.6	-5.2 2.5	-1.7 1.4	4.5 1.1	-6.3 1.6	-7.0 2.0	-3.1 2.0
		7A4	25 Est. σ_M	-6.3 2.9	-1.4 2.8	-3.4 2.8	.8 1.9	-.2 2.0	-3.2 2.5	-3.7 2.6	-3.0 2.6
		7B1	41 Est. σ_M	-2.4 1.9	-3.2 2.2	-.4 2.4	.7 1.6	.8 1.2	-3.9 2.4	-5.0 2.1	-.6 2.1
7B2		28 Est. σ_M	-.6 2.3	-3.5 1.8	-3.2 1.9	-1.6 1.2	-.5 1.3	-2.4 1.9	-1.1 1.8	1.9 1.6	

TABLE X—(Continued)

City	School	Class	Size of Sample	Mean Scores							
				Courtesy	Fair Play	Honesty	Loyalty	Mastery	Poise	Prop. Rts.	School Drive
Norwalk, Conn.	Franklin	7B3	26 Est. σ_M	-5.3 1.7	-3.8 2.2	-4.2 2.4	-3.9 1.8	3.1 1.5	-6.4 2.3	-4.3 2.2	-3.0 2.0
	Average for Franklin School		230	-1.1	-2.8	-1.2	.2	1.3	-3.2	-2.8	1.2
	Average for Norwalk		462	-1.1	-3.2	-.6	.8	.1	-4.0	-2.2	2.4
Scottsbluff, Neb.	Central	6A	41 Est. σ_M	-.9 1.5	-.1 1.4	-2.5 1.9	-.1 1.4	2.2 1.1	-1.7 1.6	-2.4 1.4	-2.8 1.6
	6B		18 Est. σ_M	-3.6 2.0	-2.4 3.2	-1.7 3.1	-1.5 2.2	4.2 1.5	-2.7 2.6	-2.9 2.8	-4.3 2.5
	East Ward	6A, B	45 Est. σ_M	.3 .7	-.6 1.6	2.2 1.7	.2 1.5	-1.3 1.2	.1 1.7	-.4 1.4	-.2 1.6
	6A, B		34 Est. σ_M	-3.6 1.6	-2.6 1.4	-.7 1.6	-3.9 1.2	.7 1.2	-3.4 2.2	-.5 1.8	-.5 1.8
	Bryant	6A	12 Est. σ_M	-2.9 2.5	-3.8 2.4	-4.4 3.2	-1.7 2.8	-3.1 2.6	-3.1 3.0	-3.7 3.2	-5.0 4.2
	Lincoln	6A, B	38 Est. σ_M	-1.2 1.7	-1.2 1.6	-2.8 1.9	.2 1.6	.7 1.1	-3.4 1.4	-.3 1.6	-.3 1.9
Average M for Scottsbluff			188	-1.5	-1.4	-1.1	-.9	1.9	-3.4	-1.3	-1.5

TESTS AND MEASUREMENTS

test slightly low in all traits except Mastery, in which they test high, and Loyalty, where they test as high or higher than the standardization group. It may again be mentioned that probably the scoring scheme employed has been unduly favorable to this probationary group, but not to the other groups.

Examining the data by traits we note that wide differences are found between schools in Courtesy, Honesty and School Drive, but relatively small differences in Mastery. In so far as the test can be tested, it is a rather interesting commentary that ordinary school children, even those who have been subjected to a special character education program, do not attain as high a score in "Mastery" or "dominance" as do delinquent boys. This may not be an entirely fair appraisal, because the trait may be sex linked, and the probationary groups were composed of boys only. There is also the possibility that "character training" produces more submissive attitudes and conduct, or that many of the boys in the probationary and parental classes had been sent there because of "masterful" conduct that would not be present in "young gentlemen."

The reason for the relatively high standing of three groups in Loyalty, and the mediocrity of the other groups in this trait, as measured, is hardly to be attributed to chance. Apparently some sort of successful tuition has had a measurable effect in the three schools, Ethical Culture, Public School 93, and Kenosha. The high correlation of this trait with School Drive may suggest a single nuclear trait characterized perhaps by conscientiousness and loyalty to self-imposed obligations. Many other group differences than those mentioned are undoubtedly significant, in view of the size of the standard errors of the means, but their causes are quite unknown to the writers.

As part of an undertaking to establish a social studies curriculum, the tuitional differences undoubtedly present in the different schools studied, which have led to the significant trait differences as measured, should be a prime object of investigation. Such a study lies beyond the purpose of the authors, which has been to provide a measuring instrument which would serve as a tool for

SUMMARY

the revelation of character differences. The test apparently accomplishes this purpose when groups are compared. It is clear that it does not do so for individuals. It is hoped that use will be made of the present instrument, without improvement, in studying group differences in character traits, and finally that the instrument may be improved to the point where it may be used in the future in the measurement of individual differences.

Three suggestions, the outgrowth of the later work with the test, may be added to the six of Section IV for improving the administration and scoring of the free association test so as to yield a more reliable and a more diagnostic instrument. From an analytical study of the eight traits, which is not as yet completed, it is apparent that they are far from being independent one from another. Because of this, (suggestion 7) great improvement could be brought about by a new selection of traits so that a small number, perhaps not over three or four genuinely independent characters, could be dealt with. (8) An increase of the number of stimulus words, followed by a selection of those found to be the more diagnostic, should yield a test with even fewer than 200 words, which would, nevertheless, be much more accurate in its measurement. (9) Having data upon the significance of the many responses to many words, it should be possible to build up a semi-controlled association test involving the more significant stimulus words and the more significant only of the responses thereto, which test would have less irrelevant material in it than does a strictly free association test.

SECTION IX

SUMMARY

This investigation has been an attempt to develop practical techniques for measuring social attitudes or character. The association test was selected for experiment because it seemed unlikely to be amenable to falsification through conscious attempts of persons being examined to modify their scores in directions which might for any reason be thought advantageous. Although the techniques of the investigation would be equally appropriate with more ma-

TESTS AND MEASUREMENTS

ture individuals, this study has concerned itself with children of approximately seventh-grade maturity. Eight social attitudes or character traits commonly named as desirable were chosen for investigation.

For the classes involved in the derivation of the scoring scheme, the members of each class who stood high and the members who stood low were defined for each trait by means of ratings by teachers and anonymous reports by fellow pupils. For each trait, responses typical of the high group were positively credited, and those typical of the low group negatively credited in building up the scoring scheme. These scoring keys, developed from the responses of seventh-grade pupils in North Carolina and the District of Columbia, were then tested on a research group yielding coefficients of reliability from .46 on Courtesy to .74 on Regard for Property Rights, and correlations with teacher and pupil ratings from .02 on Mastery to .28 on Honesty, and upon a second group consisting of junior-high-school pupils in Minnesota, yielding coefficients of reliability ranging from .00 in "Mastery" to .57 in "Honesty in School Work" and in "School Drive," and correlations with pupil and teacher ratings ranging from —.10 in "Loyalty" to .18 in "Courtesy." These results were distinctly discouraging, although numerous points had come to light at which definite improvements in procedures and techniques could easily be made.

PARSIMONY IN THE JUDGMENT AND MEASUREMENT OF CHARACTER TRAITS

TRUMAN L. KELLEY

The material studied and reported in the chapter upon Measurement of Certain Traits of Character provided data for an analytical study of character traits. We may be interested in the following issues:

(a) To what extent are the eight traits as judged independent one of another, or if not independent, to what extent can independent traits, eight or less in number, be used to characterize the same children, and how are these new traits related to the eight upon which judgments were made?

(b) We may ask questions of similar import dealing with the eight character measures resulting from scoring the association test.

(c) Thirdly, we may investigate to see if there are similarities in the analyses of the judgment measures and of the association test measures. We should expect similarity if our techniques are sound, for the scoring procedure was built up so that the association test scores would reflect the judgment ratings in so far as possible in the case of the group employed in determining the scoring scheme. However, since there are halo effects in the judgment data (see p. 434 for still another effect) and correlation between errors in the scoring of the association test,—the same words, not different ones, being scored in eight different ways to get the trait measures,—it is clear that there is much need to investigate the adequacy of the procedures involved which were designed to allow properly for these systematic errors in the two types of measures.

(d) Finally, we may ask if there is evidence suggesting that a smaller number of new traits be employed,—whether judgments constitute the original data or association test scores.

The data herein used are those for the reserve group of 67 and the Elgin group of 311 because none of it was involved in determining the scoring scheme.

TESTS AND MEASUREMENTS

As a preliminary approach let us note a few correlations corrected for attenuation for the group of 311. These are given in the upper left portion of Table XII. Property Rights and Courtesy have an intercorrelation of 1.007 (the excess above 1.00 may readily be attributed to errors in the measures involved), thereby indicating practical identity of these two traits as judged. Let us note the correlation of each with each other variable. We have the following:

	Fair Play	Honesty	Loyalty	Mastery	Poise	School Drive
Courtesy	.958	.910	.905	— .429	.808	.886
Regard for Prop. Rts.	.940	.924	.919	— .311	.805	.932

As a further check upon the possible value of the techniques employed, the mean scores of pupils in certain parental and probat onary schools of New York City were compared with the mean scores of pupils in the Ethical Culture School and in other schools in which there was reason to believe the pupils possessed better than average social attitudes. The differences between the mean scores in the two extreme types of schools were several times the standard errors of these means, indicating that these differences were statistically significant, except in the trait called "Mastery," in which the classes from the probat onary schools showed higher scores than did the classes from schools in which definite character education had been given. Whether this trait called "mastery" or "dominance" is really a desirable social attitude is perhaps not finally determined, but this was the only one of the eight traits in which "bad" pupils equalled or surpassed "good" pupils.

A study of the tables of mean scores of pupils of various classes in conjunction with the character education programs employed and the presumptive levels of the groups in the traits reveals evidence of real group differences, but even more strikingly such study shows that wide group differences in these character traits are the exception. Perhaps this is equivalent to saying that most character education programs are rather ineffectual in modifying children. So broad a conclusion as this does, of course, call for further

MEASUREMENT OF CHARACTER TRAITS

careful investigation. The evidence supports the belief that some programs have greater effect than others, but we believe it hardly warranted at this time to attempt to list the character education programs upon the basis of their effect upon each of the eight character traits. If any such attempt were made, we should be called upon to conclude that the life of the delinquent boy constitutes a good training in Mastery or dominance. We are somewhat hesitant to do this, or to draw conclusions concerning character education programs as related to the superior showing of certain "good" groups in Courtesy, Honesty, Poise, Regard for Property Rights, and School Drive. Though certain conclusions of this sort are probably warranted from the data, our knowledge of the need for improvement of the instrument, and our awareness that even such differences as are revealed are small in comparison with the variability of the groups studied, inclines us to urge further study before conclusions of broad import with reference to character education programs are drawn.

The authors are disappointed, of course, that the weights they assigned to the responses to the words in their association test do not yield measures of social attitudes of sufficient reliability and validity to justify their use in measuring the social characters of individual pupils. They believe, however, that both the reliability and validity of such measures can in future studies be greatly improved, and they recommend the use of the test as it stands for further experimental work in measuring gross differences between junior-high-school classes in the following traits: Courtesy, School Drive, Regard for Property Rights, Honesty in School Work, and Fair Play in Games and Contests. The association test, even as it was scored in this study, is capable of distinguishing among junior-high-school groups that differ in these traits, and it should be possible in experimental situations to measure by means of it the relative amounts of change produced in the character traits of classes by different types and amounts of training, environment, and the like, provided these amounts of change are of the order of, say, one-half a class standard deviation.

Both by the correlation of the two variables and the correlation

TESTS AND MEASUREMENTS

of each with each other variable, we are warranted in believing that instead of two measures we have for all practical purposes but one. This seems surprising in view of the definitions of the traits as provided for the judges. Courtesy is a trait whose extremes are characterized in the following words:

(Positive)

Considerate, courteous, polite, gracious, diplomatic, refined, well-bred, tactful, gentlemanly (ladylike), chivalrous, urbane.

(Negative)

Inconsiderate, discourteous, impolite, rude, unmannerly, undiplomatic, crude, ill-bred, tactless, ungentelemanly (unladylike), barbarous, boorish, rustic.

and Regard for Property Rights is a trait whose extremes are characterized as follows:

(Positive)

Always recognizes an owner's right to control and to protect his own property.
Asks permission before he borrows anything; returns things promptly and in good condition.
Avoids injury to property of others, but tries to make good any injury done accidentally.
Tries hard to find the owner in order to return anything he has found.

(Negative)

Takes things he wants, regardless of who owns them.
Forgets to return and does not take care of things he has borrowed.
Deliberately defaces and injures the property of the school and of others.
Runs away and denies responsibility when he injures someone's property.
Keeps anything he finds, even when he knows to whom it belongs.

No single judge of the many who made judgments volunteered the idea that the two traits were the same. Nevertheless they turn out to be. Where different terms are employed there seems to be established the conviction that different things must be involved, though at times, as here, we have evidence indicating this not to be the case in any important sense. Whatever difference there is is trivial in comparison with the sameness that is present. If Courtesy and Regard for Property Rights are not psychologically two traits but one, we are immediately led to

MEASUREMENT OF CHARACTER TRAITS

ask if some of the other measures may not merely be dependent functions. It is this more general problem which is here attacked. The endeavor is to see the extent to which some smaller number of variables than eight are adequate to account for all the sorts of variability represented by the initial eight variables.

The method of analysis used is a modification of that presented by Dr. Harold Hotelling.¹ The purpose to be accomplished and the steps for doing it may be summarized as follows: (a) Original data are given showing the correlation between m variables (in the first study here reported eight judgment character scores for the reserve group of 67). Considering each variable equally important, we assign a variance of 1.00 to each. The situation then has a total variance of m . (b) Various combinations of these variables, or if each variable represents a dimension in m -dimensional space, various rotations of axes, may be investigated. No more and no less, except in exceptional cases, than m new variables, or axes, are needed to exactly characterize or duplicate the initial situation in its sundry sorts of variability. Variability in the direction of the old axes is as given in the second column below and along the new axes in some manner, say as given in the 4th column.

TABLE I
ILLUSTRATION (HYPOTHETICAL) FOR FIVE VARIABLES

Initial Five Variables		Five Derived Variables	
Variables	Variance*	Variables	Variance*
1	1.00	a	1.30
2	1.00	b	.50
3	1.00	c	2.15
4	1.00	d	1.00
5	1.00	e	.05
Sum = 5.00		Sum = 5.00	

*The "variance" is the squared standard deviation.

Variable *a* is some weighted linear combination of variables 1, 2, 3, 4, and 5, and variable *b* some linear differently weighted

¹Hotelling, Harold: "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology*, Vol. XXIV, Nos. 6 and 7, September and October, 1933.

TESTS AND MEASUREMENTS

combination of the same initial variables, etc. No information is given by variables a, b, c, d , and e above that to be found in variables 1, 2, 3, 4, and 5. It is merely differently organized. Various conditions which may be imposed on the new variables and which it would seem will greatly facilitate thinking are, first, that a, b, c, d , and e be independent of each other. The initial variables are of course correlated by some complex pattern measured by ten intercorrelation coefficients, but quite beyond the mind to picture. The table of intercorrelations and variances for the initial variables is as follows:

TABLE II
INTERCORRELATIONS AND VARIANCES

		Initial Variables				
		1	2	3	4	5
Initial Variables	1	1.00				
	2	r_{12}	1.00			
	3	r_{13}	r_{23}	1.00		
	4	r_{14}	r_{24}	r_{34}	1.00	
	5	r_{15}	r_{25}	r_{35}	r_{45}	1.00
Variances		1.00	1.00	1.00	1.00	1.00

The table for the derived variables is

TABLE III
* INTERCORRELATIONS AND VARIANCES

		Derived Variables				
		a	b	c	d	e
Derived Variables	a	1.00				
	b	.00	1.00			
	c	.00	.00	1.00		
	d	.00	.00	.00	1.00	
	e	.00	.00	.00	.00	1.00
Variances		1.30	.50	2.15	1.00	.05

MEASUREMENT OF CHARACTER TRAITS

Clearly, all of one's thinking processes become simplified if variables *a*, *b*, *c*, *d*, and *e* are used instead of variables 1, 2, 3, 4, and 5.

However, the imposition of the condition of independence between variables is not sufficient to uniquely determine them, for an infinite number of new systems, such as that of variables *a*, *b*, *c*, *d*, and *e*, can be derived from variables 1, 2, 3, 4, and 5, all showing independence between measures. We may impose another condition. What this other condition should be may be consequent to psychological considerations or even physiological or social, but until such are made explicit and their value shown, a very real consideration from the standpoint of simplicity of thought is the principle of parsimony. Thus if our new variables are A, B, C, D, and E, such that A accounts for as great a proportion of the initial variance as possible, B for as great an amount of what is left as possible, etc., we have a system which still further simplifies thinking. The table is as follows:

TABLE IV
INTERCORRELATIONS AND VARIANCES

	DERIVED VARIABLES				
	A	B	C	D	E
A	1.00				
B	.00	1.00			
C	.00	.00	1.00		
D	.00	.00	.00	1.00	
E	.00	.00	.00	.00	1.00
Variances	3.00	1.25	.50	.20	.05

Whereas the initial variables 1, 2, 3, 4, and 5, each included because it presumably measured some important aspect, were roughly equally important on the basis of the judgments that led to their inclusion and equally potent in terms of the complex

TESTS AND MEASUREMENTS

situation represented by the initial five variables, the derived variables A, B, C, D, and E are not equally potent—variable A accounting for 60 times as much of the variance of the total situation as variable E. We may believe that variable E is not of the same importance as variable A in describing such individual differences as one is in general concerned with describing. This is, of course, an assumption, for no evidence is produced showing that variables A, B, C, D, and E do more richly characterize individual differences in situations going beyond those represented by the initial variables 1, 2, 3, 4 and 5.

The immediate aim of this study has been to use judgment measures of traits and to derive uncorrelated measures of the sort A, B, C, D, and E, wherein E can be neglected with trivial loss (in the hypothetical illustration, .01 of the total variance being thereby unaccounted for), D with a slightly greater loss (.04 of the total variance being thereby unaccounted for), etc.

The method of principal components referred to has been followed with the following changes: instead of starting with a correlation table such as Table I, we start with the following table in which r_1, r_2, \dots are reliability coefficients.

TABLE V
INTERCORRELATIONS AND VARIANCES

	VARIABLES				
	1	2	3	4	5
1.....	r_1				
2.....	r_{12}	r_2			
3.....	r_{13}	r_{23}	r_3		
4.....	r_{14}	r_{24}	r_{34}	r_4	
5.....	r_{15}	r_{25}	r_{35}	r_{45}	r_5
Variances.....	r_1	r_2	r_3	r_4	r_5
Total variance to be accounted for = Sum = $(r_1 + r_2 + r_3 + r_4 + r_5) < 5.00$					

In other words, we do not attempt to account for such portion of the variance as is due to chance, but only for such portion as is due to differences which are presumably real.

MEASUREMENT OF CHARACTER TRAITS

When dealing with the intercorrelations between association test scores, we found that coefficients corrected for attenuation were systematically greater than 1. This was clearly due to correlation between errors in scoring, since the same pupil response items, not different ones, were scored to obtain each of the eight character measures. We removed this systematic tendency¹ before writing down the initial table of intercorrelations to be analyzed. This added step has undoubtedly increased the size of the probable errors of our component variances and of the constants entering into the equations for obtaining the components, but we do not believe it has had any other effect. Unfortunately the probable errors mentioned are unknown.

For the two groups the basic correlations are as given in Tables VI, VII, VIII, IX, X, XI, and XII accompanying. Variables 1 to 8 inclusive refer to "traits as judged" by teachers and fellow students, and variables 11 to 18 inclusive refer to the same traits and in the same order "as scored" by means of the free association test.

A comparison of the correlations of Table VI with those of Table VII will show many similarities and perhaps some discrepancies, but as a comparison of the correlated complex of traits as determined by the two methods of measurement will be much simpler when based upon the components found, and the weightings of the traits yielding these components, no immediate examination is made of the correlation coefficients taken in pairs from these tables.

An examination of the correlation coefficients of Tables VII and X reveals that many coefficients corrected for attenuation by the formula

$$r_{\infty} = \frac{r_{12}}{\sqrt{r_1} \sqrt{r_2}}$$

will exceed 1.00. Further, they will exceed 1.00 by amounts not attributable to chance. This is undoubtedly due to correlation between errors in the scores as derived. A correction for this has

¹See Appendix A at the end of this paper.

TESTS AND MEASUREMENTS

been worked out as given in Appendix A. Its application yielded Table VIII from Table VII, and the lower right quarter of Table XI from Table X.

TABLE VI
CORRELATIONS FOR RESERVE GROUP OF 67

		Traits as Judged							
		1	2	3	4	5	6	7	8
Courtesy.....	1	.891							
Fair Play.....	2	.598	.834						
Honesty.....	3	.592	.688	.879					
Loyalty.....	4	.486	.565	.644	.879				
Mastery.....	5	-.338	-.105	-.064	-.107	.878			
Poise.....	6	.460	.571	.617	.644	-.036	.837		
Prop. Rts.....	7	.776	.677	.641	.527	-.249	.500	.880	
Sch. Drive.....	8	.647	.611	.777	.574	-.017	.545	.711	.918

TABLE VII
CORRELATIONS FOR RESERVE GROUP OF 67

		Traits as Scored							
		11	12	13	14	15	16	17	18
Courtesy.....	11	.462							
Fair Play.....	12	.693	.619						
Honesty.....	13	.620	.698	.525					
Loyalty.....	14	.567	.663	.671	.556				
Mastery.....	15	-.107	.038	.086	.153	.435			
Poise.....	16	.661	.803	.700	.698	.216	.642		
Prop. Rts.....	17	.783	.745	.791	.634	-.138	.645	.739	
Sch. Drive.....	18	.778	.584	.724	.569	-.137	.510	.832	.718

MEASUREMENT OF CHARACTER TRAITS

TABLE VIII

CORRELATIONS, RESERVE GROUP OF 67, INTERCORRELATIONS
MODIFIED TO ALLOW FOR CORRELATION
BETWEEN ERRORS IN SCORING

		Traits as Scored							
		11	12	13	14	15	16	17	18
Courtesy	11	462							
Fair Play	12	429	619						
Honesty	13	350	467	525					
Loyalty	14	330	474	430	556				
Mastery	15	002	065	107	150	435			
Poise	16	417	602	430	481	141	642		
Prop Rts	17	526	580	630	496	102	534	739	
Sch Drive	18	542	428	520	393	136	322	686	.718

TABLE IX

CORRELATIONS, GROUP OF 311

		Traits as Judged							
		1	2	3	4	5	6	7	8
Courtesy	1	821							
Fair Play	2	798	846						
Honesty	3	739	729	802					
Loyalty	4	696	758	689	.720				
Mastery	5	319	304	186	194	673			
Poise	6	629	642	652	637	188	738		
Prop Rts	7	835	791	757	713	234	632	.837	
Sch Drive	8	758	722	782	659	183	650	.804	.890

TABLE X

CORRELATIONS, GROUP OF 311

		Traits as Scored							
		11	12	13	14	15	16	17	18
Courtesy	11	280							
Fair Play	12	633	.237						
Honesty	13	652	.698	380					
Loyalty	14	580	664	665	286				
Mastery	15	307	173	125	184	.245			
Poise	16	584	726	726	678	.038	425		
Prop Rts	17	750	.682	.725	604	199	.597	.300	
Sch. Drive.	18	.704	570	674	.567	157	.543	.711	.363

TABLE XI
GROUP OF 31I, INTERCORRELATIONS MODIFIED TO ALLOW FOR
CORRELATIONS BETWEEN ERRORS IN SCORING

		Traits as Judged								Traits as Scored							
		1	2	3	4	5	6	7	8	11	12	13	14	15	16	17	18
Courtesy	1	.821															
Fair Play	2	.798	.846														
Honesty	3	.739	.729	.802													
Loyalty	4	.696	.758	.689	.720												
Mastery	5	-.319	-.304	-.186	-.194	.673											
Poise	6	.629	.642	.652	.637	-.188	.738										
Prop. Rts.	7	.835	.791	.757	.713	-.234	.632	.837									
Sch. Drive	8	.758	.722	.782	.659	-.183	.650	.804	.890								
Courtesy	11	.134	.098	.142	.060	.014	.101	.108	.136	.280							
Fair Play	12	-.029	-.004	.036	.010	.118	.048	-.054	.022	.150	.237						
Honesty	13	.056	.058	.121	.064	.136	.157	.026	.075	.273	.320	.380					
Loyalty	14	-.049	-.033	.028	.019	.070	.079	-.049	-.009	.195	.253	.317	.286				
Mastery	15	-.048	-.075	-.042	-.058	.055	-.008	-.070	-.015	-.113	.035	-.064	-.046	.246			
Poise	16	.012	.009	.080	.069	.129	.101	.003	.020	.190	.339	.370	.362	.031	.425		
Prop. Rts.	17	.079	.044	.086	.055	.086	.119	.064	.092	.251	.182	.321	.210	-.045	.260	.300	
Sch. Drive	18	.124	.092	.154	.070	.062	.136	.103	.156	.266	.187	.316	.261	-.086	.184	.283	.363

MEASUREMENT OF CHARACTER TRAITS

TABLE XII
GROUP OF 311, SAME RELATIONSHIPS AS SHOWN IN TABLE XI EXCEPT THAT
CORRELATIONS ARE CORRECTED FOR ATTENUATION

	Traits as Judged								Traits as Scored								
	1	2	3	4	5	6	7	8	11	12	13	14	15	16	17	18	
Courtesy 1	1.000																
Fair Play 2	.958	1.000															
Honesty 3	.910	.885	1.000														
Loyalty 4	.905	.971	.906	1.000													
Mastery 5	-.429	-.403	-.253	-.278	1.000												
Poise 6	.808	.813	.847	.873	-.266	1.000											
Prop. Rts. 7	1.007	.940	.924	.919	-.311	.805	1.000										
Sch. Drive 8	.886	.833	.926	.823	-.237	.803	.932	1.000									
Courtesy 11	.280	.200	.299	.133	.032	.222	.223	.272	1.000								
Fair Play 12	-.066	-.008	.083	.023	.296	.114	-.120	.048	.581	1.000							
Honesty 13	.100	.102	.219	.122	.270	.297	.047	.129	.836	1.065	1.000						
Loyalty 14	-.102	-.066	.059	.042	.160	.171	-.100	-.018	.689	.970	.962	1.000					
Mastery 15	-.106	-.164	-.096	-.138	.134	-.019	-.155	-.033	-.432	.144	-.210	-.172	1.000				
Poise 16	.020	.016	.137	.124	.241	.180	.004	.033	.550	1.066	.920	1.038	.095	1.000			
Prop. Rts. 17	.159	.088	.174	.118	.127	.254	.128	.179	.837	.683	.950	.715	-.165	.727	1.000		
Sch. Drive 18	.228	.166	.286	.137	.126	.263	.187	.275	.834	.638	.850	.810	-.287	.467	.857	1.000	

TESTS AND MEASUREMENTS

Tables VI, VIII, IX, lower right hand quarter of XI, and XII have been analyzed to determine the major components in each. The reason Table XII instead of Table XI was studied is that the smallness of the reliability coefficients for the test scores as compared with the judgment ratings would of itself necessitate a much greater weighting in an analysis into components of the traits as judged in comparison with the traits as scored by the association test. Greater weightings would obscure the object of the analysis which is to ascertain similarities in the ratings in the traits as judged and the same traits as scored. Therefore in this instance a table of coefficients corrected for attenuation constituted the basic table. The issue of similarity between traits as judged and the same traits as scored is a theoretical one so that the correction for attenuation is warranted. But such correction has not been followed in connection with the other tables and should not be followed if the practical issue of how to combine measures and get component scores is being considered. It seems to the writer that the type of table which in general should be analyzed is Table V wherein initial variables are roughly judged to be of equal importance. A very simple modification of the entries of such a table may be made if the initial variables are considered of unequal importance. If the judged weights are w_1, w_2, w_3, \dots all that is necessary is to multiply each reliability coefficient by the square of the corresponding weight and each correlation coefficient by the product of the two corresponding weights.

Tables XIII, XIV, XV, XVI, and XVII herewith give the results of the analyses of the various tables. Since in every instance the greater part of the variance is represented in two or three components, and since in the tables which have involved correction for correlation between errors and for attenuation, large chance errors in the analyses must be present, these have been carried to the determination of three or four components only, except in the case of Table VI where the analysis has been carried to the limit just to show what one complete analysis looks like and to emphasize the negligible importance of the last several components. The writer attaches no special significance, other than the fact that

MEASUREMENT OF CHARACTER TRAITS

TABLE XIII
RESERVE GROUP OF 67, JUDGMENTS

k values	Components								Remaining components	All
	I	II	III	IV	V	VI	VII	VIII		
	4 591	1 007	534	271	226	200	106	060	0004	6 996
	Weights of variables yielding these components								*Residual σ 's	
Courtesy 1	800	— 318	222	064	.229	004	— 207	— 034	014	
Fair Play 2	.804	052	028	313	— 254	114	004	— 093	010	
Honesty 3	856	166	011	— 160	— 217	— 052	— 130	133	014	
Loyalty 4	758	148	— 408	— 137	124	284	009	— 011	000	
Mastery 5	— 200	864	252	093	120	061	— 036	020	000	
Poise 6	728	223	— 365	150	114	— 296	026	— 002	.0001	
Prop. Rts. 7	843	— 188	242	105	104	054	186	126	.000	
Sch. Drive 8	846	150	250	— 286	— 016	— 096	097	— 126	000	

*Standard deviation of the proportion of the initial variables unaccounted for by Components I, II, III, IV, V, VI, VII and VIII.

TESTS AND MEASUREMENTS

they are small, to the weightings for components IV, V, VI, VII, and VIII. Certain imaginary variances are occasionally to be found in these tables, e. g., the residual standard deviation for Poise, Table XIII, is given as $\sqrt{-.0001}$. The value should be .00. The imaginary value is due either to computational error or to an error in the obtained reliability coefficients which have led to a coefficient of correlation corrected for attenuation which is greater than 1.00.

In the accompanying tables variables are numbered in the same order as in the preceding tables.

The k-values give the variances of the various components and their sum equals the sum of the entries in the diagonal of the correlation table and is the total variance represented by the complex situation.

TABLE XIV
RESERVE GROUP OF 67, SCORES

		Components			
		I	II	III	Remaining components
					All
k values		3.521	.719	.281	.175
					4.696
		Weights of variables yielding these components			*Residual σ 's
Courtesy	11	.619	-.115	-.009	.257
Fair Play	12	.727	.164	-.213	.132
Honesty	13	.683	.026	.201	.130
Loyalty	14	.638	.230	.067	.304
Mastery	15	.041	.580	.294	.100
Poise	16	.692	.312	-.256	.024
Prop. Rts.	17	.849	-.195	.004	$\sqrt{-.0192}$
Sch. Drive	18	.732	-.391	.195	$\sqrt{-.0091}$

*Standard deviation of the proportion of the initial variables unaccounted for by Components I, II and III.

MEASUREMENT OF CHARACTER TRAITS

TABLE XV
GROUP OF 311, JUDGMENTS

		Components				
		I	II	III	Remaining components	All
k values		5.211	.625	.212	.279	6.327
		Weights of variables yielding these components			*Residual σ 's	
Courtesy	1	.887	-.078	.107	.127	
Fair Play	2	.887	-.067	-.048	.229	
Honesty	3	.857	.115	.031	.231	
Loyalty	4	.811	.058	-.159	.184	
Mastery	5	-.305	.756	.005	.090	
Poise	6	.760	.077	-.333	.210	
Prop. Rts.	7	.898	.039	.147	.090	
Sch. Drive	8	.877	.137	.198	.251	

*Standard deviation of the proportion of the initial variables unaccounted for by Components I, II and III.

TABLE XVI
GROUP OF 311, SCORES

		Components				
		I	II	III	Remaining components	All
k values		1.927	.411	.164	.015	2.517
		Weights of variables yielding these components			*Residual σ 's	
Courtesy	11	.439	-.236	.042	.173	
Fair Play	12	.461	.205	-.031	$\sqrt{-.0188}$	
Honesty	13	.630	.005	-.011	$\sqrt{-.0167}$	
Loyalty	14	.521	.082	-.109	$\sqrt{-.0039}$	
Mastery	15	-.086	.389	.295	.010	
Poise	16	.586	.294	-.119	$\sqrt{-.0190}$	
Prop. Rts.	17	.495	-.109	.148	.146	
Sch. Drive	18	.507	-.240	.161	.149	

*Standard deviation of the proportion of the initial variables unaccounted for by Components I, II and III.

TESTS AND MEASUREMENTS

TABLE XVII
GROUP OF 311, JUDGMENTS AND SCORES
COEFFICIENTS CORRECTED FOR ATTENUATION

		Components				
		I	II	III	Remaining components	All components
k values		7.064	5.521	1.438	1.977	16.00
		Weights of variables yielding these components			*Residual σ 's	
Courtesy	1	.822	-.540	-.015	.183	
Fair Play	2	.801	-.539	.018	.259	
Honesty	3	.857	-.424	.093	.277	
Loyalty	4	.810	-.493	.131	.288	
Mastery	5	-.184	.446	.276	.831	
Poise	6	.821	-.342	.177	.421	
Prop. Rts.	7	.805	-.565	-.006	.179	
Sch. Drive	8	.806	-.454	.090	.369	
Courtesy	11	.640	.540	-.423	.347	
Fair Play	12	.473	.827	.353	$\sqrt{-.032}$	
Honesty	13	.646	.804	.009	$\sqrt{-.064}$	
Loyalty	14	.492	.841	.055	.217	
Mastery	15	-.209	-.046	.856	.472	
Poise	16	.504	.758	.373	.182	
Prop. Rts.	17	.601	.660	-.159	.421	
Sch. Drive	18	.632	.577	-.305	.418	

*Standard deviation of the proportion of the initial variables unaccounted for by Components I, II and III.

If the variances of the components are represented as percentages of the total variance, we have, for Table XIII:

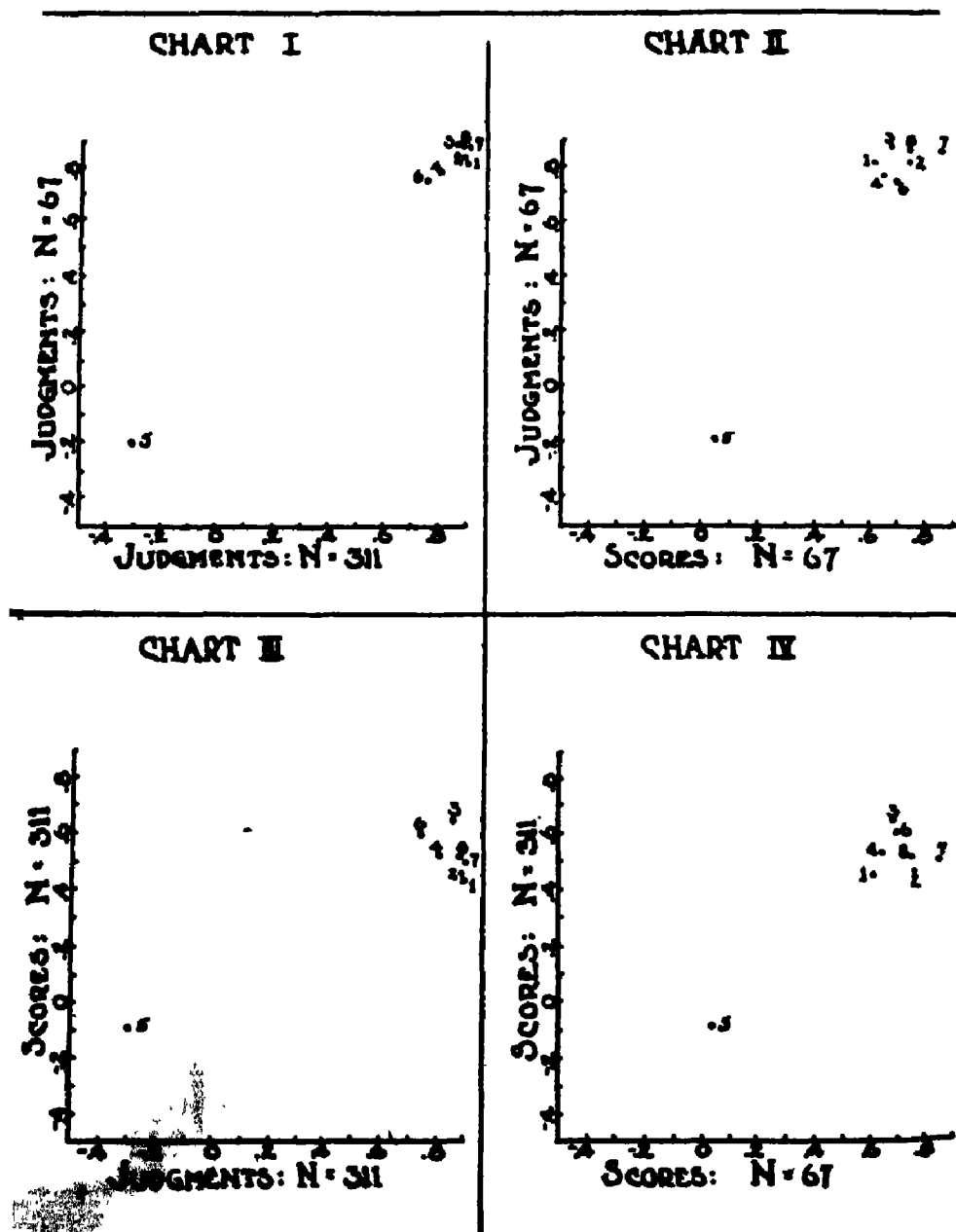
Components.....	I	II	III	IV	V	VI	VII	VIII
Variances as percentages of total.....	66	14	8	4	3+	3-	1+	1-

MEASUREMENT OF CHARACTER TRAITS

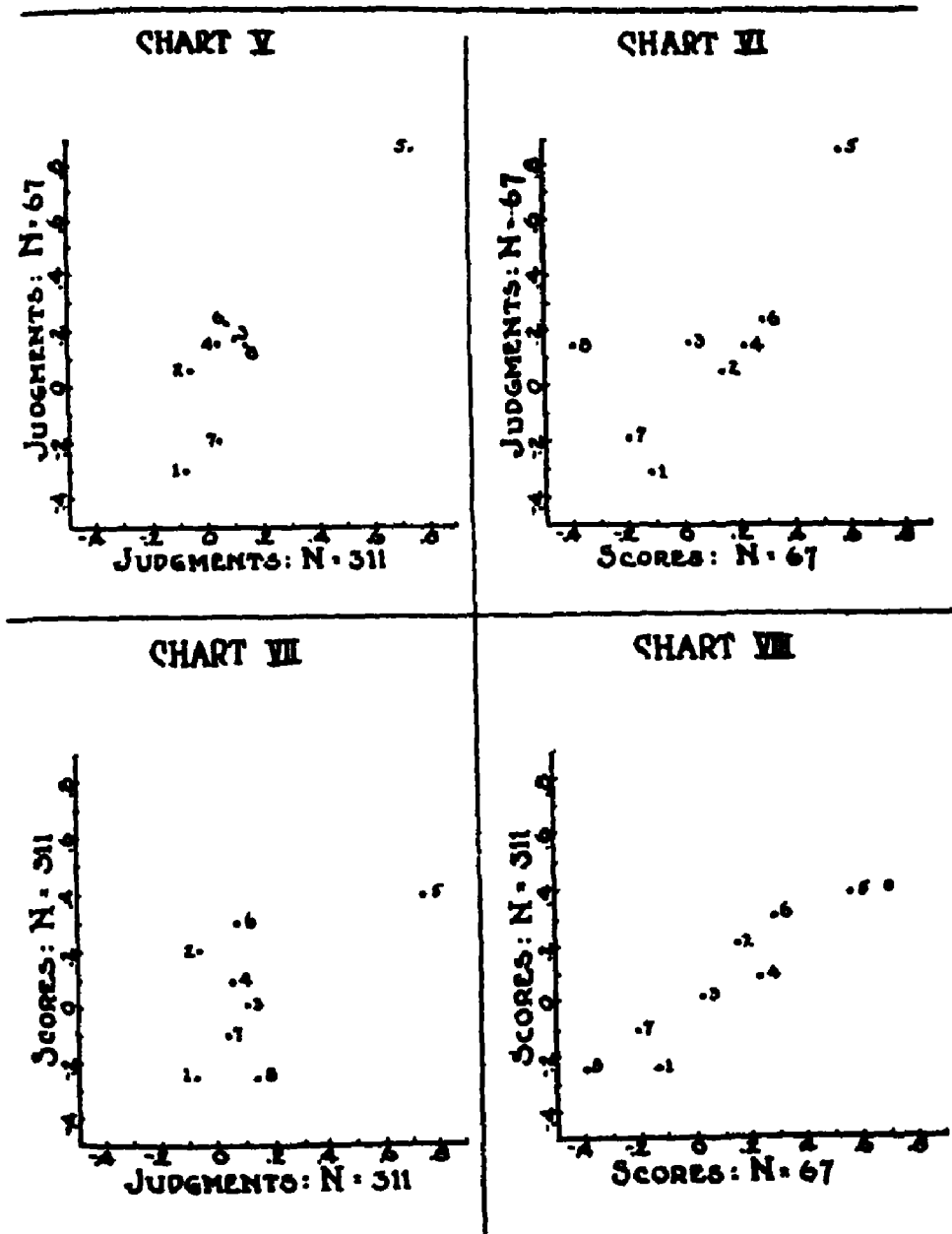
We see that a replacing of the initial variables by two new ones, component I and component II, will yield 80 per cent as much information about the variability of the pupils as do their ratings upon the entire eight traits. Of course appropriate names and descriptions for the new variables should be devised and ratings secured in harmony therewith before we can actually profit by using these components. This might prove a difficult undertaking but it surely seems worth attempting. Probably three components would suffice to tell all that is worth recording, for there is certainly much error in the weightings of the minor components, components IV, V, VI, VII, and VIII probably having negligible real meaning, that is, they are probably not much other than chance.

To get component I from the eight judgment scores (these having been reduced to standard measures), we employ the equation $x_1 = .800z_1 + .804z_2 + .856z_3 + .758z_4 - .200z_5 + .728z_6 + .843z_7 + .846z_8$. x_1 are scores whose mean is zero and whose standard deviation equals $\sqrt{k_1}$. Similarly we can obtain components II, III, etc., using the weights of the columns of Table XIII for these components. A study of the weights of component I leads us to consider that this component is a general average of all the traits except that the trait Mastery is omitted, and except that Poise and Loyalty are slightly underweighted in this average. If we now examine component I for the judgment scores for the group of 311, we see from Table XV that it may be characterized in the same words. The correlations between the weightings of the first component for the group of 67 and the first component for the group of 311 are shown graphically in Chart I wherein the numbers written near each point refer to the traits as numbered. Clearly nearly the same thing is represented by component I of Table XIII as is shown by component I of Table XV. Examining component I for the judgment scores and component I for the association test scores leads us to compare column I, Table XIII with column I, Table XIV, column I, Table XV, with column I,

WEIGHTS OF VARIABLES ENTERING INTO THE FIRST COMPONENTS OF VARIABLES IN EACH ANALYSIS



WEIGHTS OF VARIABLES ENTERING INTO THE SECOND COMPONENTS & VARIABLES IN EACH ANALYSIS



WEIGHTS OF VARIABLES ENTERING INTO THE THIRD COMPONENTS OF VARIABLES IN EACH ANALYSIS

CHART IX

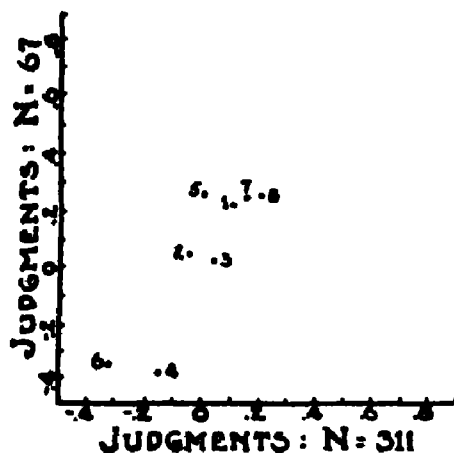


CHART X

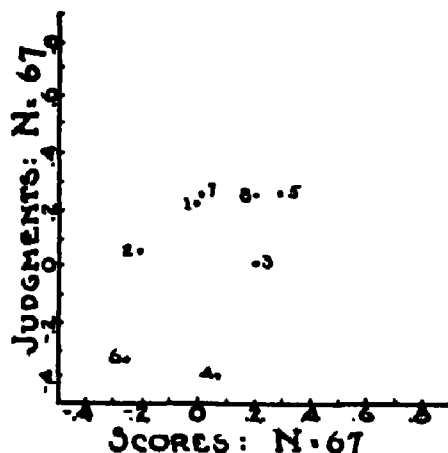


CHART XI

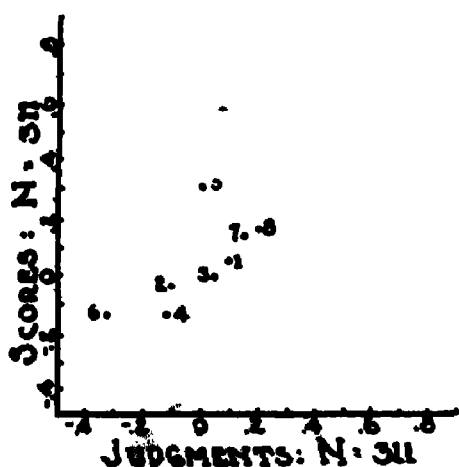
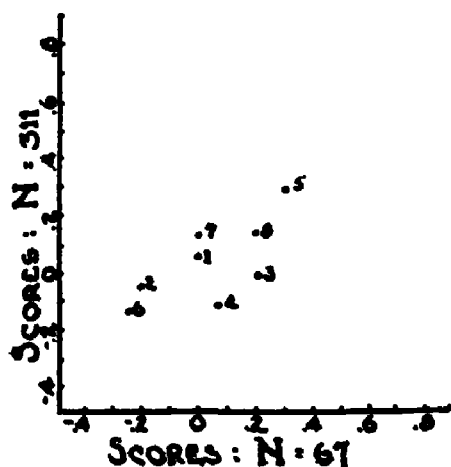
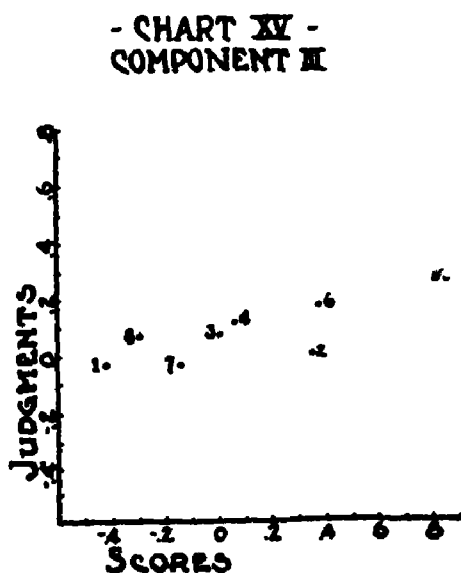
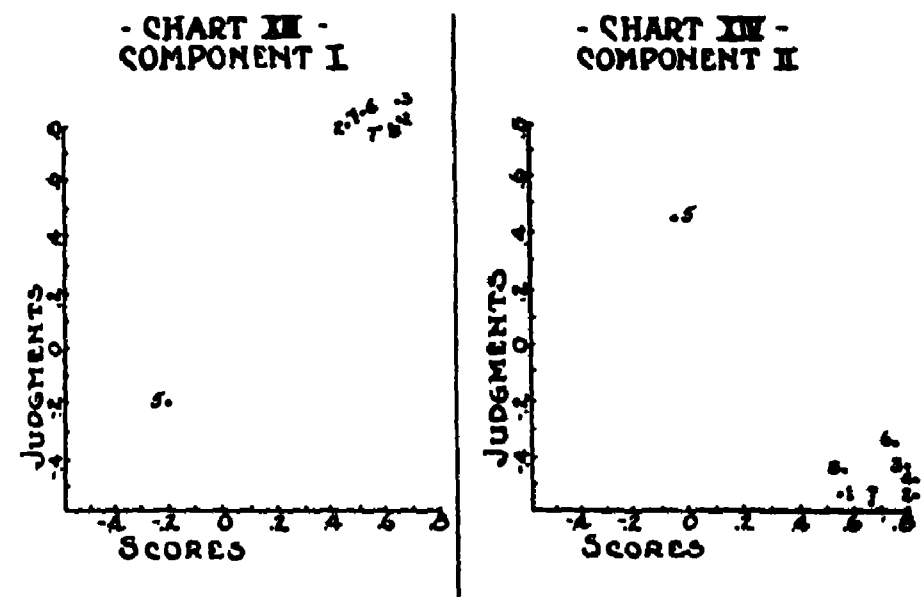


CHART XII



16 VARIABLES N = 311
CORRELATIONS CORRECTED FOR ATTENUATION



TESTS AND MEASUREMENTS

Table XVI, and column I, Table XIV, with column I, Table XVI. Gathering these four sources of data together yields the following:

WEIGHTINGS YIELDING THE FIRST COMPONENTS BY EACH OF THE FOUR ANALYSES

Variables	Judgments		Scores		Variables
	N=67	N=311	N=67	N=311	
Courtesy 1.....	.80	.89	.62	.44	11
Fair Play 2.....	.80	.89	.73	.46	12
Honesty 3.....	.86	.86	.68	.63	13
Loyalty 4.....	.76	.81	.64	.52	14
Mastery 5.....	-.20	-.30	.04	-.09	15
Poise 6.....	.73	.76	.69	.59	16
Prop. Rts. 7.....	.84	.90	.85	.50	17
Sch. Drive 8.....	.85	.88	.73	.51	18

The relationships are shown graphically in Charts I, II, III, and IV. It is clear that component I from traits as judged and component I from traits as scored are very nearly the same thing. We need a name if merely for the purpose of the present discussion, for the thing nuclear to these four component I measures. Certain names which have been suggested for this trait are "social mindedness," "social conformity," "good citizenship," "virtue." Since it seems to be the epitome of all those virtues cherished by the American forefathers, and since we do not wish to be involved in the necessity of proving it identical with some trait already named, let us call it "Puritania."

Remembering that the chief consideration in the choosing of the eight traits by a consensus of competent people was that they were separately important, we see that the outstanding characteristic of Puritania is its sweep or generality. The certainty with which the first component can be isolated is a function of its magnitude and the ratio of its variance to that of the next smaller component. For all four 8-variable analyses the ratio of the variances of the first component to the second is greater than that of the second to the third, and thus the certainty with which we can speak of Puritania will be greater than that with which we can speak of any second or third component variable. However, these latter seem far from meaningless as will be apparent from a

MEASUREMENT OF CHARACTER TRAITS

study of the weightings of the initial variables which yield them.

The study of the weightings of component II, Tables XIII and XV shows in each instance a predominance of the judged trait Mastery.

Tables XIII, XIV, XV, and XVI, from which the accompanying weightings have been drawn, and Charts V, VI, VII, and VIII are to be examined in order to make a study of the weightings of the second components as given by the judgment and the score variables.

WEIGHTINGS YIELDING THE SECOND COMPONENTS
BY EACH OF THE FOUR ANALYSES

Variables	Judgments		Scores		Variables
	N=67	N=311	N=67	N=311	
Courtesy 1.....	-.32	-.08	-.12	-.24	11
Fair Play 2.....	.05	-.07	.16	.20	12
Honesty 3.....	.17	.12	.03	.00	13
Loyalty 4.....	.15	.06	.23	.08	14
Mastery 5.....	.86	.76	.58	.39	15
Poise 6.....	.22	.08	.31	.29	16
Prop. Rts. 7.....	-.19	.04	-.20	-.11	17
Sch. Drive 8.....	.15	.14	-.39	-.24	18

Such a study indicates less similarity in the four second components than was the case in the four first components. It may be hazardous to attempt to name a nuclear trait underlying these four components, though there is something heavily weighted with Mastery, slightly with Poise and Loyalty, and slightly and negatively with Courtesy and Regard for Property Rights, which is determinative of the nature or direction of this component. The slight element of lack of consideration and the slight elements of self-confidence and loyalty should be added to the concept of dominance well to characterize the nucleus of the four second components. They do, however, differ sufficiently one from another to leave considerable doubt as to the fitness of any name which might be chosen. "Rugged Individualism" is one suggestion offered to the writer and it will serve as a name for the present discussion, though upon study it may be found to be very similar to Allport's Ascendency or Bernreuter's Dominance. Its essential characteristics are, first, complete independence of Puritania, and,

TESTS AND MEASUREMENTS

second, an importance in accounting for the variability represented by the eight initial variables of about one in seven. The reader must firmly fix in mind that there is no antagonism between these first and second component variables. Possession of a large amount of the one presages nothing as to a probable amount of the other. Thinking in terms of these variables is peculiarly simple because while thinking of one the other can be entirely neglected. If with reference to a seventh grade child knowledge of the one is at hand and later knowledge of the other becomes available, it is completely additive information calling for no discounting because in part already known. It is as though the child formerly one-dimensional now becomes two-dimensional. Such other dimensions of the child's mental life as may be independent of these two (perhaps one or more types of ability) would be similarly contributive of entirely new information, but what these are and whether they exist or not cannot be foretold previous to an experimental study of them in connection with these two.

A study of the third components resulting from the four analyses gives weightings of similarly named initial variables which are definitely and positively correlated, but the correlation is so low that doubt is cast upon any name which might be ascribed to all four third components.

We will not give a name to this nucleus. The accompanying weightings suggest a certain instability factor,—lack of Poise, of Loyalty, and of Fair Play, but accompanied by a certain vigor represented by the presence of Mastery and School Drive.

WEIGHTINGS YIELDING THE THIRD COMPONENTS BY EACH OF THE FOUR ANALYSES

Variables	Judgments		Scores		Variables
	N=67	N=311	N=67	N=311	
Courtesy 1.....	.22	.11	-.01	.04	11
Fair Play 2.....	.03	-.05	-.21	-.03	12
Honesty 3.....	.01	.03	.20	-.01	13
Loyalty 4.....	-.41	-.16	.07	-.11	14
Mastery 5.....	.25	.00	.29	.30	15
Poise 6.....	-.36	-.33	-.26	-.12	16
Prop. Rtn. 7.....	.24	.15	.00	.15	17
Sch. Drive 8.....	.25	.20	.20	.16	18

MEASUREMENT OF CHARACTER TRAITS

Possibly one end of this trait is represented by schizophrenia and the other by a well integrated personality. This is merely a suggestion and in no sense a conclusion.

The analyses thus far described have been of judgments on the one hand and of scores on the other. In the case of the group of 311 we have all 16 measures for each pupil and have correlated every one with every other one, yielding a table of intercorrelations, Table XI, which could be analyzed to yield components running through all 16 measures. Such an analysis might be expected to yield more conclusive components than the separate analyses of the judgment scores and of the association test scores. However, an examination of Table XI shows such a large excess in the reliability coefficients of the judgments in comparison with those of the scores that any such analysis would of necessity be mainly determined by the judgments, and the weightings of the 16 variables to yield components would be affected in a rather obscure manner by the differences in reliability of the two sorts of measures. Accordingly, instead of analyzing the coefficients of Table XI, we have corrected all the coefficients of this table for attenuation giving Table XII, and have analyzed these relationships. This again is not very satisfactory because correction for attenuation when initial reliabilities lie between .25 and .43, as is the case with the test scores, introduces a chance element which it is impossible to eliminate, or, with any formula at hand, to estimate. We are thus, even with this 16 variable problem, forced to restrict our positive conclusions to things revealed by the first two or three components only.

Remembering that based upon an earlier sample the scoring scheme of each response item for each trait was determined by the consensus of the judgments upon the pupil for the same trait, we should not anticipate any difference except that of reliability between Courtesy as judged and Courtesy as scored. We should, therefore, expect the weights for variables 1 and 11, 2 and 12, . . . 8 and 18 to correlate very highly for component I, again for component II, etc. In fact we find this to be the case for component I, as is clear from a study of Table XVII and Chart XIII,

TESTS AND MEASUREMENTS

but it is not the case for component II, the reasons for which we will shortly consider. Component I is apparently Puritania and nothing else. There is nothing of the unexpected in this finding and we will immediately pass on to a consideration of component II.

This component from 16 variables and the sample of 311 bears no clear relationship to component II of any of the four preceding analyses. This lack of agreement is altogether surprising and calls for careful investigation. It being impossible to investigate it at present by experimental means, we must search for logical reasons for the discrepancy. Component II is obtained by weighting positively all the judgments and negatively all the scores with one reversal, that in the case of Mastery. Now clearly there can be nothing in the psychology of children which would lead to such an outcome. Children can scarcely be courteous when being judged and discourteous when being tested by the association test. The cause must lie somewhere in the process of recording measures. Further, all judgment scores involved in the intercorrelations between judgments are identical with the same scores when involved in the correlations with the association test scores, so we should not look for the cause of the peculiarity in any of the intercorrelations. The only measures that are left to be considered are the reliability coefficients, and here we may note that there may readily be an influence which may affect judgment reliabilities quite differently from test score reliabilities. If the half scores which led to the reliability coefficients for the judgments correlate spuriously highly, there would be no carry over to the test score reliability coefficients, and similarly no systematic error in the test score reliability coefficients would carry over to the judgment reliabilities. No large source of error in the reliability coefficients of the test scores suggests itself, but such a source for the judgments is immediately suggested to one familiar with the mental labor, uncertainty, and sense of needing support which characterize the process of passing judgment as to traits of fellow beings. The judge, and we need not say the incompetent judge, faced with the request to appraise a fellow being along a line with which he

MEASUREMENT OF CHARACTER TRAITS

supposedly is familiar, will commonly satisfy the task by accepting hearsay, rumor, or reputation. The second judge does the same thing and in doing it accepts the same hearsay, rumor, and reputation. Thus there comes to be high agreement between judges, but low agreement with the trait actually possessed by the fellow being. The tendency is a commonplace experience in adult circles. Who has not been guilty of passing on an appraisal that is in no sense independent? Surely children are equally likely to form likes and dislikes and to make judgments of fellow classmates upon the basis of hearsay and not of independent knowledge. Though we consider this inevitable where independent knowledge is lacking, nevertheless it results in the making of judgments which are not independent in the sense demanded if a true reliability coefficient is to be determined.

The writer early in this study was surprised but gratified at the high reliability coefficients found for the judgment measures. He now concludes that these are untrue, being spuriously high, and that the excess of the measured judgment reliability coefficient over its real reliability plus a certain chance element is the main cause of the second component as found in the case of the 16 variable problem. A second, though less important cause, is apparently Rugged Individualism as shown by the weightings of Mastery. Rugged Individualism is not a large enough factor to dominate or determine the direction of the second component axis, but it is large enough to show its influence here, and, because not fully satisfied here, again in later components.

An important consequence of the conclusion that the obtained reliability coefficients are spuriously high is that no interpretation of this second component based upon the psychology of school children need be sought, since the component is to be interpreted as a function of the psychology of the judges. Further, the very practical conclusion is drawn that judgments are not nearly as reliable as had been estimated in connection with these data. We may believe that in general judgments of fellow beings are not nearly as reliable as obtained reliability coefficients would suggest.

If we are right in attributing this second component, which the

TESTS AND MEASUREMENTS

reader will note is very substantial in size, its variance being 5.52 out of a total of 16, to collusion, however unintentional, between judges, we may roughly estimate what the reliability of these judgments would be if the influence of rumor and hearsay could be eliminated.

If we estimate the variance of the second component which is not due to hearsay as 2.52, which in the light of the other components is probably too large rather than too small, there remains a variance of 3.00 to be attributed to hearsay. Dividing this equally among eight variables yields .375 per variable. Thus, for example, in the case of Courtesy in place of the diagonal value

1.00 (which equals $\frac{\text{reliability as obtained}}{\sqrt{.821} \sqrt{.821}}$) we have $(1.00 - .375)$

or .625 (which equals $\frac{\text{reliability as it should be}}{\sqrt{.821} \sqrt{.821}}$). The "reliability

as it should be" to give this answer is .513. Roughly this reliability based upon a composite of teachers' and pupils' judgments is as reliable as the average of four teachers' judgments. Using the Spearman-Brown formula to obtain the reliability of a single teacher's judgment of Courtesy, we obtain .21. This is very low but not out of harmony with estimates of teacher ability to judge scholastic achievement, which estimates would indicate a reliability in judging "reading ability," "arithmetical ability," etc., of between .3 and .4. It is only necessary to assume that teachers can judge of reading and arithmetical ability somewhat more competently than they can of Courtesy, Mastery, Fair Play, etc. The indicated low reliability of teachers' judgments is a handicap to experimental study utilizing them as criteria measures. It does, however, emphasize the need of objective and reliable measures of character traits, if it is possible to obtain them. The low reliabilities of the scores from the association test as given suffer less in comparison with teachers' judgments than would be the case had teachers' judgments reliabilities of .4, .5, or .6. If an association

MEASUREMENT OF CHARACTER TRAITS

test can be built up and scored yielding reliabilities of character traits of better than .5 per grade group, it will definitely serve a need not now met by any measures and it can serve as a basis for studies of character and for studies whose aim is the derivation of a still better measuring instrument.

Only a word or two need be said about the third component from the 16 variable problem. If our interpretation is correct, the hearsay nature of the second component has disrupted the analysis if taken as reflecting psychological relationships in children. The third component, which we might anticipate would be Rugged Individualism, does show heavy weighting of Mastery as judged, but no correspondingly heavy weighting of Mastery as scored. Since component II expressing opposition in weights of traits as judged from the same traits as scored is only partially expressed so far as Mastery is concerned ($-.05$ and $.44$ in contrast with, say, Honesty $.83$ and $-.42$), a part of the essential opposition seems to be carried over, so far as Mastery is concerned, to component III, and perhaps again to component IV, for an examination of the residual standard deviation column indicates that Mastery as scored will be the main ingredient of this fourth component, and Mastery as judged a much smaller portion of it. Let us, therefore, make no further attempt to interpret the third, the fourth, or higher numbered components. The surprising feature of this 16 variable analysis is the great size of the second component and its probable cause in factors outside of the children being measured.

The conclusion of this study, that two or three components contain practically all the information to be found in the original eight, may seem contrary to "common sense." The situation in its simplest case can be reduced to two dimensions. Consider a right triangle with hypotenuse c and sides a and b . We have $a^2 + b^2 = c^2$. The quantities a^2 and b^2 correspond to the variances of components and c^2 to the total variance. If $c^2 = 1$, and $a = .95$, we obtain $a^2 = .9025$; $b^2 = .0975$; and $b = .31$. Thus b has a standard deviation of sufficient size to be far from negligible. The component b may be a variable which quite justifies a separate label and

TESTS AND MEASUREMENTS

a separate concept. We can perhaps say that roughly its right to a separate concept is represented by the size of its standard deviation, b/c , but that its importance in explaining the gross situation is represented by the size of its variance, b^2/c^2 . We may readily believe that Courtesy and Honesty are different, though believing at the same time that the importance of that in respect to which they are different is very small with respect to the importance of that with respect to which they are the same. Thinking in terms of components involves thinking in terms of (1) variances of (2) measures which are independent of each other.

It is believed that both (1) and (2) will greatly contribute to psychological analysis, though it is not claimed or implied that there are not additional requirements and aids to sound thinking in this field.

SUMMARY OF RESULTS OF THIS STUDY

It may be expected that when we have measures such as Strong Interest scores derived from a single basic set of items, we will have large error correlations just as was found to be the case with the eight different scores on the association test here used. Human judgments of fellow beings may be expected to have, in addition to the commonly recognized halo effect, a large hearsay effect. The halo effect leads to spuriously high correlations between judgments of supposedly different traits, while the hearsay effect leads to spuriously high correlations between different teachers' judgments of the same trait. The indication is that, after allowance for this, the reliability of teachers' judgments of character traits is very low.

Two or three independent character trait measures are highly successful in giving the gist of the complex story told by as many as eight trait measures chosen originally because of their separate importances in connection with school situations.

The first of these measures is Puritania (a coined term) and the second Rugged Individualism.

The method of analysis into principal components, perfected

MEASUREMENT OF CHARACTER TRAITS

by Harold Hotelling, proves straightforward and powerful, and to offer promise of wide usefulness.

The association test shows promise as a means of measuring character traits, but as thus far perfected, is entirely insufficient for satisfactory individual measurement.

There is great need for objective and more accurate measures of character traits if education is to be concerned with their development or modification.

APPENDIX A

THE METHOD OF REMOVING THE EFFECT OF CORRELATION BETWEEN ERRORS IN THE ASSOCIATION TEST SCORES

Let $x_1 = x_\infty + e_1$, let $\sigma_1 = 1$

and $x_2 = x_\omega + e_2$, and $\sigma_2 = 1$

Let r'_{12} = the correlation between x_1 and x_2 before correction for correlation between errors in e_1 and e_2 , and let r_{12} = the correlation after such correction.

Let $x_{\frac{1}{I}}$ equal the score on the first half of the first test

$x_{\frac{1}{II}}$ equal the score on the second half of the first test

$x_{\frac{2}{I}}$ equal the score on the first half of the second test

$x_{\frac{2}{II}}$ equal the score on the second half of the second test

Then $x_{\frac{1}{I}} - x_{\frac{1}{II}} = e_{\frac{1}{I}} - e_{\frac{1}{II}}$

$x_{\frac{2}{I}} - x_{\frac{2}{II}} = e_{\frac{2}{I}} - e_{\frac{2}{II}}$

$r(x_{\frac{1}{I}} - x_{\frac{1}{II}})(x_{\frac{2}{I}} - x_{\frac{2}{II}}) = r(e_{\frac{1}{I}} - e_{\frac{1}{II}})(e_{\frac{2}{I}} - e_{\frac{2}{II}})$

from which

$r'_{12} = \sigma_\infty \sigma_\omega r_{\infty\omega} + \sigma_\infty \sigma_{e_2} r_{\infty e_2} + \sigma_\omega \sigma_{e_1} r_{\omega e_1} + \sigma_{e_1} \sigma_{e_2} r_{e_1 e_2}$

TESTS AND MEASUREMENTS

Assuming $r_{\infty,} = 0$ and $r_{\infty,} = 0$

$$r'_{11} = \sigma_{\infty}\sigma_{\omega}r_{\infty\omega} + \sigma_{e_1}\sigma_{e_2}r_{e_1e_2}$$

Assuming $r_{\infty,} = 0$

and $r_{\infty,} = 0$

$$\sigma'_{11} = \sigma'_{\infty} + \sigma'_{e_1}$$

$$\text{and } \sigma'_{11} = \sigma'_{\omega} + \sigma'_{e_2}$$

$$\text{and since } \sigma_{\infty} = \sigma_1 \sqrt{r_1}$$

Kelley, Stat. [166]

$$\text{and } \sigma_{\omega} = \sigma_2 \sqrt{r_2}$$

$$\sigma_{e_1} = \sqrt{1 - r_1}$$

$$\text{and } \sigma_{e_2} = \sqrt{1 - r_2}$$

$$\text{then } r'_{11} = \sqrt{r_1} \sqrt{r_2} r_{\infty\omega} + \sqrt{1 - r_1} \sqrt{1 - r_2} r_{e_1e_2}$$

Assuming $r_{e_1e_2} = 0$ and $r_{e_1e_2} = 0$

$$r_{e_1e_2} = r_{(e_1 + e_2)(e_1 + e_2)} = r_{(e_1 - e_2)(e_1 - e_2)}$$

and therefore

$$r_{\infty\omega} = \frac{r'_{11} - r_{(e_1 - e_2)(e_1 - e_2)} \sqrt{1 - r_1} \sqrt{1 - r_2}}{\sqrt{r_1} \sqrt{r_2}}$$

To determine r_{12} the correlation between scores having reliabilities r_1 and r_2 respectively which would have been obtained if there had been no correlation between errors, use the usual formula for the relation between true and obtained coefficients, which gives:

$$r_{12} = r_{\infty\omega} \sqrt{r_1} \sqrt{r_2} = r'_{12} - r_{(x_1 - x_{e_1})(x_2 - x_{e_2})} \sqrt{1 - r_1} \sqrt{1 - r_2}$$

CHEATING IN THE CLASSROOM, WITH EMPHASIS ON THE INFLUENCE OF FRIENDS

A. M. JORDAN

University of North Carolina

The problem of how to encourage children to form habits regarded as socially desirable is an ever present one. Among these desirable habits is that of common honesty; ideally, individuals should be honest under any and all circumstances. In certain classroom situations, however, the lack of honesty is very evident. A study of cheating by children in the classroom and of the causes therefor seems likely to throw light on the larger problem. It is such an investigation that is here reported.

THE PROCEDURE

In carrying out our plans it seemed too slow a procedure to wait until children showed evidences of dishonesty in the ordinary classroom activities. Furthermore, to prove that children cheat in the usual tests in the classroom is always difficult and uncertain. It seemed better in this investigation, therefore, to make use of some of the more artificial techniques which had been developed by former investigators in the field.¹ Among the techniques available that called the I. E. R. (Institute of Educational Research) seemed most suitable for our purposes.²

The I. E. R. technique consists in administering two tests so similar in nature that, if a class should take the two tests on two mornings near in time to each other, the scores would be practically the same. When the first of this pair of similar tests is given, an answer sheet is placed underneath the test with the admonition to the child not to look at the answers until the time is up. When the test is finished, the child is asked to draw out his answer sheet from under his test and to correct the test he has just taken, placing a "c" before the correct answer and an "x" before the

¹See May and Hartshorne, *Studies in Deceit*, Book 1, 1928, pp. 28-103, for a discussion of these techniques.

²We were given permission to mimeograph these tests by May and Hartshorne.

TESTS AND MEASUREMENTS

wrong answer. Within a few days the second of this pair of tests is given in a manner differing from the foregoing in that in the first instance the teacher was absent from the classroom and the investigator turned his back, thus giving the children a chance to add, if they chose, scores they had not actually made. On the second occasion, no answers are given, and the teacher is in the room assisting the investigator to make sure that the children receive no assistance whatever.

There are then, two situations presented to the children: (a) one in which they have an opportunity for cheating and (b) another in which there is no opportunity for cheating. Couple these facts with the fact that the two tests are equivalent forms and it is seen that the scores obtained in Situation A minus those earned in Situation B give a measure of cheating. There were three such sets of tests used, (1) arithmetic problems, (2) completion of sentences, and (3) information.

In February and March of 1930 these three sets of tests were administered to the children in Grades IV, V, VI, and VII in one of our North Carolina cities. Nine grades were involved in the study: one fourth grade, four fifth grades, two sixth grades, and two seventh grades—276 children in all. These grades were located in two schools with four in one school and five in the other. In March of the following year this testing program was repeated in order to get a measure of the change in cheating habits during the intervening period of one year.

During this year several types of evidence were collected:

(1) Censuses were taken of friends at three-month intervals, four censuses in all.

(2) The investigator visited each of these classes from five to nine times to investigate the behavior of the children, both in the class and on the playground. Each visit was from two to three hours in length. The teacher of the class furnished the investigator with a seating plan by which each child could be easily located. Notes concerning the behavior of the children were entered immediately or shortly after the observation was made. Ratings

CHEATING IN THE CLASSROOM

of cooperativeness and of general good behavior were made of each child from these personal observations.

(3) Ratings of the child's cooperativeness and general classroom behavior were secured from each classroom teacher.

(4) Ratings of cooperativeness and general behavior were also secured from the physical education teacher in four classes.

(5) The Sims test of economic and cultural level was administered in December, 1930. This test consists of a series of questions such as, "Does your family own an auto which is not a truck?" "Where do you regularly spend your summers?" "How many magazines are regularly taken in your home?" which tend to throw light on the socio-economic level of the individual questioned.

(6) The occupation of the father or mother was kept as a record by the school. This information was entered into our records.

(7) The scores of the Stanford Achievement Tests were also obtained. The school gives these tests each year as a partial indication of the right of a child to progress to the next grade.

(8) The grade location and the age of each child was recorded also.

With these facts collected the question naturally arises as to what was their influence upon or relation to the matter of cheating.

AMOUNT OF CHEATING ON TESTS

Under the conditions of this test when every child has an opportunity to cheat, there is likely to be much more cheating than might be anticipated. Table I sets forth the percentages of cheating in each of the tests used, and for each grade. If we consider the cheating on the arithmetic test only, we see that the percentage of pupils cheating varies from 43 per cent in Grade V B-1 of the first school to 85 per cent in Grade V B-1 of the second school. In School I there is a tendency for cheating to increase with the advance of pupils through the grades. The percentages of cheating in arithmetic run in round numbers 62, 43, 73, 83. Note that Grade V B-1 has a percentage of cheating far below that of the other grades. In attempting an explanation the principal of this school confirmed a suspicion which the experimenter had also en-

TESTS AND MEASUREMENTS

tertained, that this class was one of the most indolent groups of children found in the whole school system. On account of these traits of character these pupils were really not enough interested in the tests to want to cheat. On the other hand, VI B-2 was one of the most ambitious classes visited. The members of this class enjoyed test comparisons with other classes in any test or in any other form of competition because they surpassed all other classes in almost every comparison. One hypothesis which might be

TABLE I
SCHOOL I

AMOUNT OF CHEATING BY GRADE AND ON EACH TEST IN THE TWO SCHOOLS STUDIED
N = Number in Class. C = Number Cheating. P = Percent Cheating.

	Grade IV, B-3			Grade V, B-1		
	Arith.	Comp.	Inf.	Arith.	Comp.	Inf.
<i>First Year:</i>						
N.....	34	36	36	28	29	29
C.....	21	11	6	11	10	9
P.....	61.8	30.6	16.7	42.9	34.5	31.0
	Grade V, B-2			Grade VI, B-2		
	Arith.	Comp.	Inf.	Arith.	Comp.	Inf.
N.....	33	34	34	29	28	28
C.....	24	16	12	24	5	11
P.....	72.7	47.1	35.3	82.8	17.9	39.3
	Grade IV, B-3			Grade V, B-1		
	Arith.	Comp.	Inf.	Arith.	Comp.	Inf.
<i>Second Year:</i>						
N.....	30	30	30	19	20	20
C.....	16	4	5	13	13	13
P.....	53.3	13.3	16.7	68.4	65.0	65.0
	Grade V, B-2					
	Arith.	Comp.	Inf.			
N.....	31	30	30			
C.....	22	5	6			
P.....	71.0	16.7	20.0			

AMOUNT OF CHEATING ON TESTS

TABLE I—Continued

SCHOOL II

	Grade V, B-1			Grade V, B-2			Grade VI, B-1		
<i>First Year:</i>	Arith.	Comp.	Inf.	Arith.	Comp.	Inf.	Arith.	Comp.	Inf.
N.....	27	33	33	35	36	36	34	33	33
C.....	23	22	23	25	22	25	22	18	17
P.....	85.2	66.7	69.7	71.4	61.1	69.4	64.7	54.5	51.5
	Grade VII, B-1			Grade VII, B-2					
N.....	28	26	26	26	27	26			
C.....	19	9	6	16	12	8			
P.....	67.9	34.6	23.1	61.5	44.4	30.8			
	Grade V, B-1			Grade V, B-2			Grade VI, B-1		
<i>Second Year:</i>									
N.....	21	20	21	22	24	23	20	18	18
C.....	15	11	14	16	13	14	8	6	4
P.....	71.4	55.0	66.7	72.7	54.2	60.9	40.0	33.3	22.2
	Grade VII, B-1			Grade VII, B-2					
N.....	20	19	17	17	15	17			
C.....	17	8	7	10	3	9			
P.....	85.0	42.1	41.2	58.8	20.0	52.9			

Comparative Findings by May and Hartshorne¹

	V, B-1	V, B-2	VI, B-1	VI, B-2	VII, B-1	VII, B-2
Public School (Suburban community, 10,000).						
Per cent cheating, population A, all tests.....	75	69	66	44	21	34
Public School (Mid-western city, 200,000)						
Per cent cheating, population B, all tests.....	41	57	25	37	31	48
Public School (metropolitan area—above average).						
Per cent cheating, population C, all tests.....	54	56	54	64	66	59

(1) May and Hartshorne, *Studies in Deceit*, Book 2, p. 54 ff.

TESTS AND MEASUREMENTS

offered is that the amount of cheating may be related to the urge present. In one case, one might hypothesize, with ambition low and the urge weak, the results are 43 per cent of cheating; in the other, with ambition overpowering and the urge strong, the results are 83 per cent of cheating. In School II there is a steady decrease in cheating beyond the fifth grade.

During the period of a year there were many changes among the responses to these tests. Referring again to Table I, one observes that in the test of the second year in School I there is a decrease of cheating in the advanced section of the fourth grade in arithmetic and completion tests; the percentage on the information test remains the same over the one-year period. In Grade V B-1, there is an increase in cheating of nearly 100 per cent in all three tests. In School II changes in a year's time may be observed in five grades. In Grade V B-1, there are substantial decreases of cheating in arithmetic and completion, but only a slight change in information tests. Grade V B-2 shows only slight changes, while Grade VB-1 shows a very substantial decrease in all three tests. Note that the population of this grade has been reduced more than the average, a fact which may account for the change. The lower seventh shows a definite increase in cheating, while the upper seventh decreased in two instances and increased in the information test.

Some of these changes are due to the fact that slightly different populations are compared. All the pupils in a grade were given the first test. A year later most of the pupils had been promoted but some had remained in the same grade and still others had shifted to different sections. This left a reduced number of those who were first tested to be tested the second time. Let it be perfectly clear, however, that all records of the testing in the second year were secured from pupils who had already been tested the first year. It is for this reason that there is always a reduction in population tested the second year.

How do the cheating records of this study compare with those of different populations? Table I presents comparative data selected from the studies of May and Hartshorne. It is clear that there is a wide variation in cheating among the public school popu-

AMOUNT OF CHEATING ON TESTS

lations selected for comparison. These data were collected by means of the I. E. R. technique also, but differed in the respect that one more test (vocabulary) was used than in the study here reported. This gave the pupils one additional chance to cheat. Then, too, the May and Hartshorne figures are for those who had cheated in any of the four situations, while the North Carolina figures are cheating scores on single tests. It is thus clear that in the May and Hartshorne technique the scores would be slightly larger than in the present case. If, however, the comparison is made of cheating scores on arithmetic tests, the difference between the results of the two investigations is very slight. In the North Carolina study on five pages of names, with thirty-five names to a page, there were only six cheating on any tests who had not also cheated on arithmetic. Cheating scores on arithmetic, therefore, will be used to compare with those of the previous study.

In Grade V B-1, we find 43 and 68 per cent of the present study cheating, percentages which may be compared with 75, 41, and 54 per cent of the previous one. In grade V B-2, the scores are 71 and 73 per cent cheating in the present study in comparison with 69, 57, and 56 of the previous one. Thus of the two fifth grades the lower section is below, the advanced section above, those studied by May and Hartshorne. In Grade VI B-1 the 65 per cent of this investigation is to be compared with the 66, 25, and 54 per cents of the previous one. In VI B-2, these scores are 83 per cent in the first and 44, 37, and 64 per cent in the second. The sixth grades of the North Carolina schools thus with one exception make larger scores on cheating tests than do these other schools. Grade VII B-1 shows 68 per cent cheating, a larger number than the 21, 31, and 66 per cent of the previous investigation. In Grade VII B-2 there are 61 per cent who cheat, which may be compared with 34, 48, and 59 of the previous investigation. It seems clear that the cheating records in the present investigation show slightly less cheating than the May and Hartshorne records in Grade V, but considerably more in the upper grades. This greater amount in the upper grades looms still larger if the comparisons are made

TESTS AND MEASUREMENTS

with Populations A and B only, for with Population C there is a substantial increase of cheating in the upper grades.

The results show that the cheating in the population studied in this investigation is slightly greater than among those of a similar nature investigated by May and Hartshorne.

TABLE II
PERCENTAGE CHEATING AT EACH AGE

C. = Number cheating. N. C. = Number not cheating. Per. = Per cent cheating.

Age.....	9			10			11			12		
	C.	N. C.	Per.	C.	N. C.	Per.	C.	N. C.	Per.	C.	N. C.	Per.
School I.....	17	8	68	30	14	68	32	9	78	6	2	75
Age.....	13			14			15					
	1	3	25	4	..	100			
Age.....	9			10			11			12		
School II.....	..	1	0	18	3	86	28	5	85	28	10	74
Age.....	13			14			15					
	22	12	65	15	4	79	7	1	87			
Age.....	9			10			11			12		
Total.....	17	9	65	48	17	74	60	14	81	34	12	74
Age.....	13			14			15					
	23	15	61	19	4	83	7	1	87			
May and Hartshorne, age.....	9			10			11			12		
Population A, Per.....	45			41			19			36		
Population B, Per.....	..			41			32			33		
							34			41		
										44		

In Table II, the percentage of cheating is calculated for each age. The total scores undoubtedly give us the truer picture because of the larger numbers of cases included. Considering, therefore, the total scores, there is evident a gradual rise in the percentages cheating until the eleventh year is reached. Ages twelve and thirteen show a dropping off with another rise at the years fourteen and fifteen. At age fifteen there are not enough cases from which

RELATION OF SOCIO-ECONOMIC STATUS TO CHEATING

to make inferences, but what data there are indicate an increase of cheating at that age. We may say, in general, that there is little or no diminution of cheating with age but rather an increase.

The results from the work of May and Hartshorne¹ are so different from those here reported that they need considerable comment. In the first place, their results do not seem to be consistent. For example, in Grades V B-1, V B-2, and VI B-1, of Population A, the percentages of cheating for all pupils are 75, 69, and 66, respectively. The pupils most likely to occupy these grades are those of ages ten, eleven, twelve, and a few at age thirteen. But the percentage cheating at these ages are 41, 19, 36, and 35. All these figures are for Population A and were achieved on the I. E. R. school test. When all the members of the population are assigned to grade sections (reported as Grade V, Grades V and VI, or simply Grade VI, which were not included in the grade percentages mentioned above) the percentages run 75, 66, 50, and 44 for grades V B-1, V B-2, VI B-1, and VI B-2. Not a year of ages ten, eleven, twelve, and thirteen has a percentage as high as the lowest reported here. We have then the anomalous situation of the members of a grade cheating from 65 to 75 per cent, but the ages composing the grade cheating from 19 to 41 per cent. The computation for age was evidently not based on the same figures as that for grades. Taking these figures at face value, the pupils of the present investigation age by age cheated a very great deal more than did those who were the subjects of the former investigation.

RELATION OF SOCIO-ECONOMIC STATUS TO CHEATING

The Sims test of socio-economic status was given to all the pupils included in this study who were present on the day of testing. This test is a reliable test of cultural and economic status. Questions such as the following are asked the pupils:

1. Do you take private lessons in music?
2. How many magazines are regularly taken in your home?
3. How many rooms does your family occupy? and how many persons occupy these rooms?

¹*Op. cit.*, p. 441.

TESTS AND MEASUREMENTS

There are twenty-three such questions in all. The reliability of the test has been computed by correlating the even answers with the odd in which case the coefficient was .83; and also the scores of matched sibs, which gave a coefficient of .94. In either case the reliability is sufficiently high for our purposes.

The question before us is the effect of the cultural and economic level of the family upon the child's cheating. This problem was approached in two ways: (1) the number of points possible to score on the test was divided into three groups, 5-13, 14-26, and 27-32, and the percentage of pupils cheating was computed for each group; (2) the coefficient of correlation was computed between the scores received on the cheating tests and those on the Sims test.

Table III shows the percentage of cheating at each level, for each grade, in both of the schools. Totals are also presented. From the consideration of these latter, it is clear at the first glance that, contrary to what one might expect,¹ there is an increasing amount of cheating with increasing socio-economic level. In School I the percentage cheating at the lowest economic level is 69; at the highest 77. In School II the percentage cheating at the lowest level is 75, while 100 per cent in the highest economic level cheated. Among the grades studied those in School I who scored lowest (5-13) on the Sims test tend to increase; while in School II the tendency is in the opposite direction. In School I the middle group (14-26) also increases with the higher grades; while in School II there is no tendency in any direction. In School I the group scoring highest (27-32) increase in cheating up to Grade V, then decrease somewhat; in School II in this group there is 100 per cent. cheating in every case.

¹May and Hartshorne found that there was a negative correlation between Sims socio-economic scores and cheating in school. They report, however, in a footnote on page 214, "It might be noted that the r between Sims score and School XI is positive in School P, and only slightly negative in School C, yet when the two schools are combined in one plot the r becomes $-.490$." It is then clear that in one of the schools studied the correlation between the Sims score and cheating was positive, while in the other school the r was negative. In the present study the r was almost zero in both cases.

RELATION OF SOCIO-ECONOMIC STATUS TO CHEATING

TABLE III

PERCENTAGE OF CHEATING AT EACH LEVEL OF THE SIMS SOCIO-ECONOMIC TEST

SCHOOL I

N. = Number in class. C. = Number cheating. P. = Per cent cheating.

Grade	Sims Level	N.	C.	P.
IV, B-3.....	5-13	5	2	40.0
	14-26	16	10	62.5
	27-32	15	8	53.3
	Total	36	20	55.6
V, B-1.....	5-13	9	7	77.8
	14-26	10	5	50.0
	27-32	6	6	100.0
	Total	25	18	72.0
V, B-2.....	5-13	1	1	100.0
	14-26	12	9	75.0
	27-32	20	17	85.0
	Total	33	27	81.8
VI, B-2.....	5-13	1	1	100.0
	14-26	12	11	91.7
	27-32	16	13	81.2
	Total	29	25	86.2
Grand Total.....	..	123	90	73.2
All Grades.....	5-13	16	11	69.0
	14-26	50	35	70.0
	27-32	57	44	77.0

SCHOOL II

Grade	Sims Level	N.	C.	P.
V, B-1.....	5-13	15	14	93.3
	14-26	13	12	92.3
	27-32	1	1	100.0
	Total	29	27	93.1
V, B-2.....	5-13	9	8	88.9
	14-26	21	19	90.5
	27-32	5	5	100.0
	Total	35	32	91.4
VI, B-1.....	5-13	13	8	61.5
	14-26	15	11	73.3
	27-32	4	4	100.0
	Total	32	23	71.9

TESTS AND MEASUREMENTS

TABLE III—Continued

SCHOOL II—Continued

Grade	Sims Level	N.	C.	P.
VII, A-1.....	5-13	13	8	61.5
	14-26	9	9	100.0
	27-32	1	1	100.0
	Total	23	18	78.3
VII, A-2.....	5-13	6	4	66.7
	14-26	13	11	84.6
	27-32	3	3	100.0
	Total	22	18	81.8
Grand Total.....	..	141	118	83.7
All Grades.....	5-13	56	42	75
	14-26	71	62	87
	27-32	14	14	100
All Grades, Both Schools.....	5-13	72	53	74
	14-26	121	97	76
	27-32	72	58	81
	Total	265	208	78.4

TABLE IV

PERCENTAGE CHEATING AT EACH LEVEL OF THE SIMS SOCIO-ECONOMIC TEST
BOTH SCHOOLS COMBINED

Sims Level	IV.			V.			VI.			VII.		
	N.	C.	P.	N.	C.	P.	N.	C.	P.	N.	C.	P.
5-13.....	5	2	40	34	30	88	14	9	64	19	12	63
14-26.....	16	10	62	56	45	80	27	22	81	22	20	91
27-32.....	15	8	53	32	29	91	20	17	85	4	4	100

Table IV shows the combined results for both schools. In the group scoring from 5-13 on the Sims test Grade V has a slightly higher average on the tests of cheating. Among the group receiving 14-26 on this test there is a steady increase of cheating through the seventh grade. Except in Grade IV, the 27-32 point group show

RELATION OF SOCIO-ECONOMIC STATUS TO CHEATING

almost universal cheating. There seems to be something in the school environment that causes cheating to increase. Furthermore, those of good social background seem to cheat as much as those of poor.

Tables V, VI, VII summarize in an analytical way the cheating scores received by pupils at each of the three levels on the Sims score card on socio-economic level. It must be remembered that the higher the score on the Sims score card, the higher is the economic level and that the larger the cheating scores, the greater is the amount of cheating. The tables of distribution at each grade enable one to observe directly the amount of scatter. For example, in School I, Level I, Grade IV, two very different groups of children may be seen. Out of fifteen children, eight evidently did not cheat at all, while seven cheated a great deal—as if they, once embarked on such a cheating career, were going to cheat as much as possible. None of the distributions is symmetrical.

In School I there is a much larger proportion of pupils in Level I (27-32) than in School II, there being 55 in School I and only 13 in School II. On the other hand, in School II there are 48 in Level III (5-13) while in School I there are only 12 in this level. There is not such a great difference in the second level. These facts give us added confidence in the test, for School I is attended by many children of bankers, physicians, professors, real estate agents, while many sons and daughters of operatives of textile and tobacco factories attend School II. When Sims levels are compared, it is found that cheating scores are higher in School II at all levels in all grades. One must distinguish sharply between cheating scores and the percentage cheating. Cheating scores refer to the amount of difference between two equivalent forms of a test, one given with answers; the other, without. The percentage cheating refers to the proportion of students out of the total number tested who avail themselves of the opportunity to cheat. Compare, if you will, Level I in the two schools. In School I, in Grade V B-2 the score is 37.50 (Table V); in School II, the corresponding score is 45 (Table VI). With Level II, School II scores are higher wherever comparisons are possible; namely at Grades V B-1, and V -B2. As

TESTS AND MEASUREMENTS

between the levels, Table VII shows that in Grade IV, Level I shows the smallest cheating scores followed by Level III, with Level II a poor third. In Grade V, Level III has the lowest scores,

TABLE V
CHEATING AT EACH LEVEL OF SCORES OBTAINED ON SIMS SCORE CARD
SCHOOL I

Cheating Scores	LEVEL I				Cheating Scores	LEVEL II				Cheating Scores	LEVEL III			
	IV B-3	V B-1	V B-2	VI B-2		IV B-3	V B-1	V B-2	VI B-2		IV B-3	V B-1	V B-2	VI B-2
70	70	70
65	65	1	..	1	..	65
60	60	60
55	5	1	55	2	55
50	1	..	3	1	50	..	1	1	1	50
45	1	2	45	1	45	1	1
40	5	..	1	3	40	2	..	3	1	40
35	..	2	1	..	35	1	2	35
30	2	3	30	30
25	1	1	25	2	..	2	3	25	1
20	2	..	20	..	1	..	1	20	..	3
15	1	15	1	1	15	..	1
10	2	1	10	1	1	..	2	10	..	1
5	2	2	5	2	1	2	1	5	1
0	2	1	3	3	0	1	2	0	1	2
-5	1	..	1	..	-5	2	2	..	1	-5
-10	1	-10	-10
-15	-15	1	..	-15
N	15	5	19	16	N	15	10	11	11	N	4	7	..	1

MEDIAN CHEATING SCORES AT EACH SIMS LEVEL

Grades.....	Grades			
	IV, B-3	V, B-1	V, B-2	VI, B-2
Level I (27-32).....	13.75	38.75	37.50	30.00
Level II (14-26).....	28.75	10.00	28.75	22.50
Level III (5-13).....	17.50	17.50

RELATION OF SOCIO-ECONOMIC STATUS TO CHEATING

TABLE VI
CHEATING AT EACH LEVEL OF SCORES OBTAINED ON SIMS SCORE CARD
SCHOOL II

Level I						Level II					
Cheating Scores	V B-1	V B-2	VI B-1	VII B-1	VII B-2	Cheating Scores	V B-1	V B-2	VI B-1	VII B-1	VII B-2
80.....	80.....	1
75.....	75.....	1
70.....	70.....	..	2
65.....	65.....	1	..	2
60.....	..	1	1	..	1	60.....	3	1
55.....	..	1	1	55.....
50.....	50.....	2	..	2	3	..
45.....	1	45.....	1	2	2	1	2
40.....	1	40.....	2	2	2	..	2
35.....	1	35.....	..	1	1
30.....	..	1	30.....	..	1	1	1	1
25.....	..	1	1	25.....	..	5	..	1	1
20.....	20.....	..	1	2	1	1
15.....	15.....	..	2	..	1	..
10.....	1	10.....	1	..	2
5.....	1	..	5.....	1	2
0.....	0.....	..	1	1
-5.....	-5.....	1
-10.....	-10.....
N.....	1	4	4	1	3	N.....	11	18	15	9	11

LEVEL III					
Scores.....	V, B-1	V, B-2	VI, B-1 Level III	VII, B-1	VII, B-2
80.....
75.....	1
70.....	1
65.....	..	1	1
60.....	1	1
55.....	..	1	..	1	..
50.....	1	1	1
45.....	2	1	1	2	..
40.....	..	2	2
35.....	2
30.....	1	..	1	1	..
25.....	1	1	..	3	..
20.....
15.....	2	..	1
10.....	1	..	1	2	..
5.....	2	1	..	1	1
0.....	2	2	1
-5.....
-10.....
N.....	11	9	11	12	5

TESTS AND MEASUREMENTS

TABLE VI—Continued

SCHOOL II—Continued

Grades.....	Median Cheating Scores at Each Level				
	V, B-1	V, B-2	VI, B-1	VII, A-1	VII, A-2
Level I, (27-32).....	..	45.00	50.00
Level II, (14-26).....	60.85	30.00	41.25	32.50	27.50
Level III, (5-13).....	46.25	47.50	32.50	26.67	17.50

TABLE VII

CHEATING AT EACH LEVEL OF SCORES OBTAINED ON SIMS SCORE CARD

SCHOOLS I AND II

Cheating Scores..	LEVEL I Grades					Cheating Scores..	LEVEL II Grades				
	IV	V	VI	VII	Total		IV	V	VI	VII	Total
80...	80...	..	1	1
75...	75...	..	1	1
70...	70...	..	2	2
65...	65...	1	2	2	..	5
60...	..	1	1	1	3	60...	..	4	4
55...	..	6	2	..	8	55...	2	2
50...	1	3	1	..	5	50...	..	4	3	3	10
45...	1	3	4	45...	1	3	2	3	9
40...	5	4	4	..	10	40...	2	7	3	2	14
35...	..	3	..	1	4	35...	1	3	1	..	5
30...	..	3	3	..	6	30...	..	1	1	2	4
25...	..	2	2	..	4	25...	2	7	3	2	14
20...	..	2	2	20...	..	2	3	2	7
15...	1	..	1	15...	..	3	1	1	5
10...	2	..	1	1	4	10...	1	1	3	2	7
5...	2	..	2	1	5	5...	2	3	1	3	9
0...	2	4	3	..	9	0...	1	3	1	..	5
-5...	1	1	2	-5...	2	2	2	..	6
-10...	1	1	-10...
-15...	-15...	..	1	1
N.....	15	29	20	4	68	N.....	15	50	26	20	111

RELATION OF SOCIO-ECONOMIC STATUS TO CHEATING

TABLE VII
LEVEL III

Scores	Grades				Total
	IV	V	VI	VII	
80.....
75.....	..	1	1
70.....	..	1	1
65.....	..	1	1	..	2
60.....	..	2	2
55.....	..	1	..	1	2
50.....	..	2	1	..	3
45.....	1	3	2	2	8
40.....	..	2	..	2	4
35.....	2	..	2
30.....	..	1	1	1	3
25.....	1	2	..	3	6
20.....	..	3	3
15.....	..	1	2	1	4
10.....	..	2	1	2	5
5.....	1	3	..	2	6
0.....	1	2	2	3	8
-5.....
-10.....
-15.....
N.....	4	27	12	17	60

MEDIAN CHEATING SCORES AT EACH LEVEL

	Grades				Total
	IV	V	VI	VII	
Level I (27-32).....	13.75	39.17	31.67	25.00	35.00
Level II (14-26).....	28.75	38.33	28.33	30.00	31.87
Level III (5-13).....	17.50	32.50	35.00	25.83	28.33

followed by Level II and Level I in order. In Grade VI, Level II shows the lowest score, followed by Level I and Level III in order. Finally, in Grade VII, Level I shows the smallest score, followed

TESTS AND MEASUREMENTS

by Level III and Level II in order. As a whole, the largest scores fall at Level I, followed by Levels II and Levels III. This finding is diametrically opposite to that expected.

OCCUPATIONAL LEVEL AND CHEATING

The relation between *occupational level* and *cheating* was also investigated. The school had a permanent record of the occupation of each child's parents. These records were referred to Taussig's classification. Taussig divides occupations into five levels. "Group I Professional men, proprietors of large businesses, and higher executives. Group II. Commercial service, clerical service, large land owners, managerial service of a lower order than Group I, and business proprietors employing from five to ten men. Group III. Artisan proprietors, petty officials, printing trades employees, skilled laborers, with some managerial responsibility, shop owners and business proprietors employing one to five men. Group IV. Skilled laborers (with exception of printers) who work for someone else, building trades, transportation trades, manufacturing trades involving skilled labor, personal service, small shop owners doing their own work. Group V. Unskilled laborers, common laborers, helpers, hands, peddlers, varied employment, venders, unemployed, unless it represents the leisured class or retired."¹

Tables VIII, IX, and X set forth the relation between the scores on the Taussig groups and cheating, first, school by school, and second, with the results combined. There are few general trends that can be recognized in the two schools taken separately. In School I the percentage of cheating tends to increase as one considers lower Taussig groupings; School II shows the opposite effect. When the results are combined there is a slight increase of cheating down to Group III, but a slight decrease in Groups IV and V. The summary table also shows a slight increase of cheating up to the sixth grade, although the fifth and sixth are separated by only one point. The greatest amount of cheating appears in the fifth and sixth grades.

¹See Sims's modifications of Taussig's classification, *Manual of Directions for the Sims Score Card for Socio-Economic Status*, p. 9.

OCCUPATIONAL LEVEL AND CHEATING

TABLE VIII

PERCENTAGE CHEATING AT EACH OCCUPATIONAL LEVEL

SCHOOL I

N=Number in class. C=Number cheating. P=Percent cheating.

Classes	Taussig's Classification					Total
	I	II	III	IV	V	
Grade IV, B-3						
N ..	8	14	5	9		36
C.	4	8	5	6		23
P	50 0	57 1	100 0	66.7	..	63.9
Grade V, B-1						
N... .. .	3	5	8	13	1	30
C.	1	1	5	10	..	17
P	33.3	20 0	62 5	76 9	..	56.7
Grade V, B-2						
N	10	13	8	4	..	35
C.	7	8	7	3	..	25
P	70 0	61 5	87.5	75.0	..	71.4
Grade VI, B-2						
N...	7	8	6	8	..	29
C....	6	7	5	7	..	25
P	85 7	87 5	83.3	87.5	..	86.2
Total N....	28	40	27	34	1	130
Total C.....	18	24	22	26	..	90
Total P.....	64 3	60 0	81.5	76.5	..	69.2

TESTS AND MEASUREMENTS

TABLE IX

PERCENTAGE CHEATING AT EACH OCCUPATIONAL LEVEL
SCHOOL II

Classes	Tausig's Classification					Total
	I	II	III	IV	V	
Grade V, B-1						
N.....	..	5	9	17	1	32
C.....	..	4	7	15	1	27
P.....	..	80.0	77.8	88.2	100.0	84.4
Grade V, B-2						
N.....	2	11	13	6	6	38
C.....	2	10	11	5	6	34
P.....	100.0	90.9	84.6	83.3	100.0	89.5
Grade VI, B-1						
N.....	3	8	7	7	10	35
C.....	2	8	4	3	7	24
P.....	66.7	100.0	57.1	42.9	70.0	68.6
Grade VII, B-1						
N.....	1	4	5	11	5	26
C.....	1	2	4	8	3	18
P.....	100.0	50.0	80.0	72.7	60.0	69.2
Grade VII, B-2						
N.....	2	1	7	11	5	26
C.....	2	..	5	6	3	16
P.....	100.0	..	71.3	54.5	60.0	61.5
Each Class						
Total N.....	8	29	41	52	27	157
Total C.....	7	24	31	37	20	119
Total P.....	87.5	82.8	75.6	71.2	74.1	75.8

RELIABILITY AND VALIDITY OF THE TESTS USED

TABLE X

TAUSSIG'S OCCUPATIONAL CLASSIFICATION AND CHEATING

Both Schools Combined

N. = Number in class. C. = Number cheating. P. = Per cent cheating.

Classes	Taussigs' Groups					Total
	I	II	III	IV	V	
Grade IV						
N.....	8	14	5	9	..	36
C.....	4	8	5	6	..	23
P.....	50	57	100	67	..	64
Grade V						
N.....	15	34	38	40	8	135
C.....	10	23	30	33	7	103
P.....	67	68	79	83	88	76
Grade VI						
N.....	10	16	13	15	10	64
C.....	8	15	9	10	7	49
P.....	80	94	69	67	70	77
Grade VII						
N.....	3	5	12	22	10	52
C.....	3	2	9	14	6	34
P.....	100	40	75	64	60	65
Total						
N.....	36	69	68	86	28	287
C.....	25	48	53	63	20	209
P.....	69	70	78	73	71	69

RELIABILITY AND VALIDITY OF THE TESTS USED

The tests used in the investigation here reported were chosen because their reliability and validity had already been carefully determined. At the same time, it seemed useful to know how valuable they were with the population studied. Measures were at hand to compute both the reliability and the validity.

A—Reliability of Intelligence Tests

As measures of intelligence the battery of three tests used correlate as follows:

TESTS AND MEASUREMENTS

TABLE XI

RELIABILITY OF INTELLIGENCE SCORES (One year apart)

Intel. I and Intel. II.....	= .742 (N, 167)
Arith. (B) I and Arith. (B) II.....	= .763 (N, 183)
Completion (B) I and Completion (B) II.....	= .702 (N, 185)
Information (B) I and Information (B) II.....	= .457 (N, 182)

These coefficients computed between scores separated by an interval of a year are a measure both of the reliability of the tests when used as measures of intelligence and of the change which had taken place in the individual in intelligence in a year's time. May and Hartshorne report a coefficient of .87, which indicates the reliability of the tests as measures of intelligence when these measures are taken close together. If this .87 is compared with the present .74 between measures a year apart, changes in the individuals themselves are indicated.

The reliability both of the batteries (represented by "Intel. I" and "Intel. II") and of the individual tests is sufficient for our purposes.

B—Validity of Intelligence Tests

These tests of intelligence were correlated with grade placement as measured on the Stanford Achievement Test and with the ratings of cooperativeness made by the teachers and the investigator.

TABLE XII

CORRELATIONS OF INTELLIGENCE TESTS AND STANFORD
ACHIEVEMENT GRADE PLACEMENTS

Intel. I and Stanford Achievement.....	= .820 (N, 271)
Arith. Reas. (B) I and Stanford Achievement....	= .667 (N, 280)
Completion (B) I and Stanford Achievement....	= .679 (N, 280)
Information (B) I and Stanford Achievement....	= .572 (N, 279)
Intelligence II and Stanford Achievement.....	= .739 (N, 176)
Arith. Reas. (B) II and Stanford Achievement....	= .565 (N, 190)
Completion (B) II and Stanford Achievement....	= .590 (N, 189)
Information (B) II and Stanford Achievement....	= .461 (N, 186)
I = First Test. II = Second Test.	

This table shows high correlations both with the test as a whole (Intelligence I, and Intelligence II) and with the sub-tests as well.

RELIABILITY AND VALIDITY OF THE TESTS USED

It is thus seen that this test of intelligence correlates more closely with the Stanford Achievement Test given close to it than with itself given a year earlier.

One could hardly expect an intelligence test to correlate very highly with rating scores of cooperativeness because quite different traits are being measured.

TABLE XIII
CORRELATIONS OF INTELLIGENCE AND RATINGS FOR COOPERATIVENESS

Intel. I and Classroom teacher's ratings.....	= .251 (N, 279)
Intel. I and Physical education teacher's ratings..	= .424 (N, 129)
Intel. I and Investigator's ratings.....	= .171 (N, 221)
Intel. II and Classroom teacher's ratings.....	= .268 (N, 176)
Intel. II and Physical education teacher's ratings..	= .334 (N, 179)
Intel. II and Investigator's ratings.....	= .318 (N, 139)

The correlations between intelligence and ratings for cooperativeness are present, but low. By means of these correlation coefficients one is enabled to get an understanding of the sort of test that is being used under the conditions of this investigation.

C—Reliability and Validity of Cheating Tests

1. *Reliability of Cheating Tests.* The principal question before us, however, is not whether or not these tests are measures of intelligence, but whether or not they measure cheating well. This fact is doubly hard to determine because, much more than intelligence, cheating depends upon the inclination of the moment. It is for this reason that higher correlation is found among the subtests than between the results of the same test given one year apart.

TABLE XIV
RELIABILITY OF MEASURES OF CHEATING TAKEN ONE YEAR APART

First Cheating Scores (sum of three tests) with second cheating scores.....	= .480 (N, 148)
Cheating scores on two arithmetic tests 1 year apart	= .433 (N, 169)
Cheating scores on two Completion tests 1 year apart.....	= .336 (N, 169)
Cheating scores on two Information tests 1 year apart.....	= .305 (N, 168)

TESTS AND MEASUREMENTS

The highest coefficient here is only .48, but, before condemning the test as lacking in reliability, let the following table be considered:

2. *Validity of Cheating Tests.*

TABLE XV
INTERCORRELATIONS OF TESTS OF CHEATING

Cheating score Arithmetic with cheating score Completion.....	= .610 (N, 201)
Cheating score Information with cheating score Completion.....	= .570 (N, 282)
Cheating score Information with cheating score Arithmetic.....	= .539 (N, 264)

From these data it is clear that the elements of the test have "marked" correlations with each other. It may be inferred from this fact that the test, as a whole, measures cheating fairly well since there is such a consistency between the elements of the test. Precisely the same type of situation is presented to the children in the case of each test and hence intercorrelations between the subtests taken near together may be regarded as the better indicator of their reliability. The next section, as a whole, may be considered in connection with the validity of these tests of cheating.

The intercorrelations of the ratings in the present investigation are not much different in size from those found in other investigations.

The closest resemblance found is that coefficient between the rating of the investigator and that of the teacher of physical education ($r = .456$).

TABLE XVI
INTERCORRELATION OF RATINGS OF COOPERATIVENESS

Experimenter's ratings of cooperativeness with room teacher's...	= .412 (N, 235)
Physical education teacher's ratings of cooperativeness with room teacher's.....	= .354 (N, 136)
Experimenter's ratings of cooperativeness with physical education teacher's.....	= .456 (N, 130)

RELATION OF CHEATING TO OTHER VARIABLES

RELATION OF CHEATING TO OTHER VARIABLES

A—Cheating Scores of Pupils' Friends

The relation between the cheating scores of children and those of their friends constitutes the central theme of this investigation. By means of censuses of pupils' friends taken at three-month intervals over a period of twelve months, it was hoped to gain some insight into the role that friends play in cheating. Coefficients of correlation were consequently computed between the cheating scores of each individual and those of his four (or fewer) best friends. This made eight correlations in all because each pupil had two cheating scores, one on the first test, and another a year later.

TABLE XVII

COEFFICIENTS OF CORRELATION BETWEEN CHILDREN'S OWN CHEATING SCORES
AND THE CHEATING SCORES OF FRIENDS

	Census I	Census II	Census III	Census IV
Own Score (1)....	.30 (N, 215)	.282 (N, 185)	.048 (N, 155)	.098 (N, 146)
Own Score (2)....	.203 (N, 146)	.161 (N, 128)	.315 (N, 122)	.306 (N, 136)

"Own Score (1)" was given at the beginning of the year at which time the first census of friends was taken. The first (I), second (II), third (III), and fourth (IV) censuses of friends appeared at successively greater distances from the cheating test. The decreasing correlations at Censuses II, III, and IV are accounted for by changing friends and changing scores in the same friends. The second cheating score, Own Score (2), correlates more highly with the third and fourth censuses which were near in time to it. The third and fourth censuses were taken within three months of the second cheating score, just as the first and second censuses were within three months of the first cheating score. These two lines of correlation are consistent in showing "present but low" correlations between individual cheating scores and present friends' and negligible correlations between individual cheating scores and

TESTS AND MEASUREMENTS

friends' scores taken 6-9 months in either direction. If the factor of age is partialled out from the relation between a pupil's own cheating score and that of his friends, the .30 of the first census becomes .295, a negligible change.

Similar results are obtained when correlations are computed between the cheating scores of the four censuses of friends. The average of the four friends' cheating scores was used with each census.

TABLE XVIII
INTERCORRELATIONS OF FRIENDS' CHEATING SCORES¹

	II	III	IV
I.....	.32 (N, 190)	.208 (N, 157)	.158 (N, 148)
II.....		.331 (N, 141)	.249 (N, 128)
III.....			.567 (N, 133)

Censuses near together in time, I and II; II and III; III and IV, thus have much higher coefficients than those more distant in time, I and III, I and IV, and II and IV. If there is any influence among friends in the matter of cheating, it comes from those friends who are contemporaneously associated and not from those removed by a short period of time.

Another method which aids in interpreting the results is that of Pearson's coefficient of mean square contingency. By means of this twofold classification there may be computed the difference between chance resemblance and that which is present. For example, in Table XIX there are 75 cases studied which showed during one year a definite change either toward or away from cheating. The question to be answered is, "Is there a tendency for friends to change *pari passu* and in the same direction as for the individuals concerned?"

¹These findings agree closely with the r of .35 of the May and Hartsborne investigation when the time between tests was the same as own score with first or second cheating scores of friends. May and Hartsborne, *Studies in Deceit*, p. 275.

RELATION OF CHEATING TO OTHER VARIABLES

Table XIX shows that those students who increased their cheating from the first to the second test had friends little like themselves, for, out of 33 individuals who increased their cheating scores, only 17 sets of friends increased their scores, while 16 sets did not. Quite different results appear with those students who decreased their cheating, for their sets of friends decreased their cheating scores in 31 cases out of 42. The coefficient of contingency is somewhat higher than .267 indicates, since the highest possible coefficient in a twofold table is .707. The coefficient is likewise about .3 when correlation is computed between individuals and their closest friends' scores. This coefficient indicates a resemblance between the cheating of individuals and their friends' cheating scores. Quite interesting is the fact that in fifty out of sixty-three cases friends cheated when the individuals cheated.

TABLE XIX
CHANGED SCORES IN INDIVIDUALS COMPARED WITH CHANGES IN
FRIENDS' SCORES IN CHEATING

Individual	Friends		
	Increase	Decrease	Total
Increase.....	(12) 17	(21) 16	33 ..
Decrease.....	(16) 11	(26) 31	.. 42
	28	47	75

$$\chi^2 = \frac{25}{12} + \frac{25}{16} + \frac{25}{21} + \frac{25}{26}$$

Coefficient of Mean Square Contingency

$$C = \sqrt{\frac{5.79}{75 + 5.79}} = .267$$

Highest possible C with 2-fold classification = .707.

TESTS AND MEASUREMENTS

TABLE XIX—*Continued*

INDIVIDUAL'S CHEATING SCORE WITH THAT OF BEST FRIEND

Individual's Score:	Friends		
	N	C	
N.....	(10)	(23)	..
	17	16	33
C.....	(20)	(43)	63
	13	50	..
	<u>30</u>	<u>66</u>	<u>96</u>

N = Not cheating. C = Cheating.

Coefficient of Mean Square Contingency.

$$X^2 = \frac{49}{10} + \frac{49}{20} + \frac{49}{23} + \frac{49}{43} = 10.55$$

$$C = \sqrt{\frac{10.55}{96 + 10.55}} = .31$$

Correlations between Cheating Scores of an Individual and those of his best friend or friends show a present but low correlation.

CORRELATIONS

Individual cheating scores and those of his best friend..... = .194 (N, 192)
 Individual cheating scores and those of his 4 best friends..... = .172 (N, 134)

May and Hartshorne report an r of .159 between the cheating scores of an individual and those of his friend, when the friend was at school in another classroom.

B—Cheating Scores Compared with Ratings of Cooperativeness

Ratings of cooperativeness were obtained from the physical education teacher, the room teacher, and the investigator. In general, we should expect that those pupils who cooperated the most in promulgating the activities of the classroom or playground would score less on cheating than those lacking in this characteristic. And so it turns out. The negative coefficients indicate clearly that there is an inverse ratio between cheating and cooperativeness. The coefficient is not large, but is consistent in each case.

RELATION OF CHEATING TO OTHER VARIABLES

TABLE XX

CORRELATIONS BETWEEN CHEATING SCORES AND RATINGS OF COOPERATIVENESS

Cheating scores (all) I test and ratings (Experimenter).....	== -.117	(N, 210)
Cheating scores (all) II test and Ratings (Experimenter).....	== -.212	(N, 139)
Cheating scores (all) I test and Ratings (Classroom teacher)...	== -.061	(N, 266)
Cheating scores (all) II test and Ratings (Classroom teacher)	== -.162	(N, 158)
Cheating scores (all) I test and Ratings (Playground teacher)	== -.010	(N, 120)
Cheating scores (all) II test and Ratings (Playground teacher)	== -.406	(N, 74)

The sub-tests show similar results.

CORRELATIONS BETWEEN CHEATING SCORES AND RATINGS OF COOPERATIVENESS

Arith. (cheating score I) and cooperativeness (Experimenter's ratings).....	== -.090	(N, 120)
Completion (cheating score I) and cooperativeness (Experimenter's ratings).....	== -.20	(N, 122)
Information (cheating score I) and cooperativeness (Experimenter's ratings).....	== -.186	(N, 122)

Finally, when intelligence was made constant by the partial correlation technique, r between the experimenter's ratings and cheating scores was reduced from $-.20$ to $-.11$.

C—Cheating and Intelligence

Our results show a negative correlation of a more substantial nature between cheating and intelligence. In one case a coefficient is as high as $-.44$.

TABLE XXI

CORRELATIONS BETWEEN INTELLIGENCE AND CHEATING

Intelligence (B scores) I with cheating on 1st test.....	== -.392	(N, 264)
Intelligence (B scores) I with cheating on 2nd test.....	== -.295	(N, 152)
Intelligence (B scores) II with cheating on 1st test.....	== -.192	(N, 159)
Intelligence (B scores) II with cheating on 2nd test.....	== -.439	(N, 158)
Completion (B scores) I with cheating on 1st test.....	== -.337	(N, 264)
Completion (B scores) I with cheating on 2nd test.....	== -.340	(N, 154)
Information (B scores) I with cheating on 1st test.....	== -.259	(N, 264)
Information (B scores) I with cheating on 2nd test.....	== -.247	(N, 152)
Arithmetic (B scores) I with cheating on 1st test.....	== -.178	(N, 264)
Arithmetic (B scores) I with cheating on 2nd test.....	== -.184	(N, 155)

I = First Test. II = Second Test.

TESTS AND MEASUREMENTS

Here, once more, the results are consistent. Note that the correlations are the same whether the tests as a whole are included or whether only sub-tests are used. The less intelligent cheat more, and the more intelligent less, than the average. The results from May and Hartshorne's study are in general agreement with those here reported. Their coefficients between intelligence and cheating range from $-.33$ to $-.49$.

A part of these negative correlations might be attributable to the relation of both intelligence and cheating to the Sims score of socio-economic level. This last factor, therefore, was made constant by the partial correlation technique. The correlation between cheating and intelligence with Sims constant is $-.41$. Thus the $-.39$ coefficient between intelligence and cheating is increased to $-.41$ when the effect of the Sims scores was teased out.

D—Cheating and Age

From Table XXII it may be deduced that there is a slight increase of cheating with age. Except in one case, all coefficients are small but positive. Age and cheating scores on the second test show a present, but low, coefficient.

TABLE XXII
CORRELATIONS BETWEEN AGE AND CHEATING SCORES

Age and cheating scores (all) I.....	= .077 (N, 264)
Age and cheating scores II.....	= .222 (N, 158)
Age and cheating scores (Arithmetic) I.....	= .004 (N, 274)
Age and cheating scores (Arithmetic) II.....	= .134 (N, 159)
Age and cheating scores (Completion) I.....	= .159 (N, 281)
Age and cheating scores (Completion) II.....	= .352 (N, 175)
Age and cheating scores (Information) I.....	= .058 (N, 281)
Age and cheating scores (Information) II.....	= .142 (N, 176)

I—First Test, II—Second Test—one year later than the first.

Age is thus not closely related to the factor of cheating in Grades IV to VII, inclusive.¹

¹May and Hartshorne report coefficients ranging from $-.025$ to $.289$ with a composite of $+.146$ (our average is $.130$).

RELATION OF CHEATING TO OTHER VARIABLES

The correlation between cheating scores and age, with the scores on the Stanford Achievement Test made constant, resulted in a change from $-.004$ to $.019$, a negligible one.

E—*Cheating and Stanford Achievement Test Scores*

Our data show a present, but low negative, correlation between cheating scores and the grade placement as computed from the Stanford Achievement Test. To eliminate the factor of age to a very large extent, correlations were computed from subjects all of whom were in grades IV and V. It will be noted that the coefficient is slightly lowered under this treatment (from $-.248$ to $-.118$), but in the case of the second test the coefficient increases from $-.248$ to $-.355$. Those boys and girls who score high on their daily work are less likely to cheat than those who score low. Contrariwise, those low in their daily work are more likely to cheat than the others.

When the relation between the Sims score and cheating on the one hand and the Stanford Achievement score on the other is eliminated, the correlation of $-.25$ between cheating and the Stanford Achievement score becomes $-.277$ or $-.28$.

TABLE XXIII

CORRELATIONS BETWEEN CHEATING SCORES AND SCORES ON THE STANFORD ACHIEVEMENT TEST

Stanford and cheating scores I.....	= $-.248$ (N, 256)
Stanford and cheating scores II.....	= $-.314$ (N, 158)
Stanford and cheating scores I (Grades IV, V only)	= $-.118$ (N, 149)
Stanford and cheating scores II (Grades IV, V only).....	= $-.355$ (N, 111)
Stanford Achievement and Age.....	= $-.089$ (N, 289)

A very surprising effect is produced by partialing out the factor of intelligence from these relations. The correlation between Stanford and cheating scores, with intelligence constant, changes from $-.25$ to $+.14$, showing that this negative resemblance is due simply to the factor of intelligence, which correlates so highly with Stanford Achievement scores.

TESTS AND MEASUREMENTS

In the May and Hartshorne study there is no direct comparison which can be made but only an indirect one. They report an r of $-.20$ and $-.02$ between scholastic averages and cheating scores, coefficients slightly less significant than the ones reported in the present study between the Stanford Achievement test and cheating scores.

F—*Cheating and Sims Scores*

Finally, we come to the relation between socio-economic level and cheating.

TABLE XXIV
CORRELATIONS BETWEEN SIMS SCORES AND CHEATING

Sims and cheating I	=	.053 (N, 238)
Sims and cheating II	= -	.129 (N, 145)
Sims and Arithmetic (cheating) I	=	.157 (N, 243)
Sims and Completion (cheating) I	= -	.056 (N, 254)
Sims and Information (cheating) II	= -	.047
Sims and Arithmetic (cheating) II	= -	.051 (N, 161)
Sims and Completion (cheating) II	= -	.164 (N, 159)
Sims and Information (cheating) I	= -	.014 (N, 254)
Sims and Stanford Achievement	=	.286 (N, 265)
Sims and Intelligence I	=	.237 (N, 251)

These correlations range from a small positive relationship to a small negative relationship in such a way as to be convincing that in the present study there is little correlation between cheating scores and socio-economic level as measured in the Sims test.

This argument is fortified by the results obtained by separating out the common element of intelligence from this relationship. The coefficient of correlation between the Sims scores and cheating scores is raised from .05 to .156 when the factor of intelligence is made constant. There seems to be little, if any, relation between socio-economic level and cheating.

These results may be compared with a negative relationship of $-.49$ found by May and Hartshorne. They do say, however, that in one school the correlation is positive with cheating in school, but when the results were combined the negative relation results. On the other hand, they find little or no resemblance between

SUMMARY AND CONCLUSION

Sims scores and cheating scores on athletic and party tests of cheating.

SUMMARY AND CONCLUSION

Tests of cheating were administered to 275 pupils in nine different grades. The tests used were those developed by the Institute of Educational Research (Teachers College, Columbia) and modified by May and Hartshorne for use in their investigation. They were administered twice, with a year between tests. During this year (1) four censuses were made of the friends of the subjects, (2) visits to the class and playground were made by the investigator, (3) ratings of cooperativeness and general class behavior were obtained from the classroom teacher, the teacher of physical education, and the investigator, (4) the Sims test of socio-economic level was administered to all, (5) the occupation of the father and mother of the children was obtained from the school record, (6) scores of the Stanford Achievement Test were entered on the records, (7) and finally, the grade, location, and age of each child were entered on the records. Correlations were then computed between the different variables assembled, and the influence of certain variables were partialled out whenever such a procedure seemed at all promising.

(1) The amount of cheating under the conditions of the test was unexpectedly large. On one test of arithmetic reasoning the percentage of those cheating ranged from 43 to 85 per cent. There was a slight increase in the upper grades. In comparison with the results of May and Hartshorne, the figures are closely similar in Grades V B-1 and V B-2, but in Grades VI and VII there was much more cheating among the pupils in this investigation.

(2) Substantial changes in cheating scores appeared during the period of a year. The score in cheating is not nearly as steady as an Intelligence Quotient. The reliability coefficient, which approaches .75 when computation is made between tests that have been given near together, is reduced to .48 when the tests are given a year apart. Furthermore, 75 cases were found who changed definitely during the year, either from "cheating" to "not cheating"

TESTS AND MEASUREMENTS

or *vice versa*. Out of 75 definite changes discovered, 33 increased and 42 decreased in their cheating scores.

(3) *Age and cheating are not substantially related.* There is a slight increase in cheating from ages nine to eleven, then a decrease from twelve to thirteen, and finally another increase at fourteen and fifteen. The correlations between age and cheating range from .004 to .352 with an average of .130. This agrees substantially with the r of .146 obtained by May and Hartshorne.

(4) Little relation was found between socio-economic level as measured on the Sims score card and cheating. Pupils with excellent economic opportunities were just as likely to cheat as those with meager opportunities. The correlations between the Sims scores and cheating scores ranged from $-.056$ to .157 with an average of .002. Nor was there any relation shown when the socio-economic level was determined from the occupation of the parents as entered on the school record. When these occupations were classified into the five Taussig levels, the percentage cheating at each level beginning with the highest was 69, 70, 78, 73, 71. These results agree with the correlations between the Sims test and cheating scores. These results are different from those discovered by May and Hartshorne, who report a coefficient of $-.49$, although they report that of two schools studied one showed a positive and the other a negative correlation. Moreover, in their study no resemblance was discovered between the Sims scores and cheating scores on athletic and party tests.

(5) Some interesting measures of reliability and of other relationships were discovered in the present investigation.

- a. As measures of intelligence the I. E. R. tests showed a self-correlation of .74 when the tests were given a year apart.
- b. As measures of cheating these tests correlated .48 a year apart. This is not a true reliability because of substantial changes in cheating.
- c. These I. E. R. tests as measures of intelligence correlated .82 with Stanford Achievement grade level.
- d. As measures of intelligence, these tests correlated from .17 to .42 with teachers' and investigators' ratings of cooperativeness.

SUMMARY AND CONCLUSION

The inter-correlations of these ratings ran from .35 to .46.

(6) Cheating scores of individuals and of their friends correlated in a very significant way. The censuses were taken four times during the year at intervals of three months. The first test of cheating was given near the first two censuses; the second test of cheating near the second two censuses. The r between individual and friends' cheating scores for the first census was .30; for the fourth, .098. The coefficient between the individual's cheating scores on the second test and friends' scores of the first census was .20; of the last census, .31. There is thus a clear tendency toward greater resemblance in those scores taken near in time. When age is partialled out, the results are the same. Moreover, when the coefficient of mean square contingency was computed between the changes in cheating between 75 individuals and their friends the correlation was .267 (maximum twofold classification is .707). And finally, when this same coefficient was computed between an individual's cheating or not cheating and his best friend's cheating or not cheating the r was .31. This evidence suggests that there is a small but definite tendency for friends to influence each other in cheating.

(7) Cheating scores and ratings for cooperation show a negative correlation. The coefficients range from $-.01$ to $-.41$ with an average coefficient of $-.161$.

(8) Intelligence and cheating show the highest negative correlations. The coefficients of correlation range between $-.18$ and $-.44$ with an average at $-.286$. Those of greater intelligence tend to cheat less than those of less intelligence and *vice versa*. May and Hartshorne found here an r of $-.493$.

(9) Cheating also is negatively related to the Stanford Achievement Test. The coefficients in this case ranged from $-.118$ to $-.355$ with an average of $-.259$. There is thus a slight tendency for those who are well advanced in their work to cheat less and *vice versa*.

CHAPTER VI

SUMMARY AND CONCLUSION

A. C. KREY

The plan upon which the Commission conducted its Investigation set aside one major division of the work for the preparation and application of new-type tests. The preceding pages have presented a fairly complete record of the work of this division, more or less in the order in which the work was done. The conclusions to be derived from this narrative of experience in the construction and use of new-type tests in the field of the social science subjects are probably already clear to those who have read these pages with care. It is a record of high hopes, great difficulties, many disappointments, and some achievements. These are here briefly summarized.

In retrospect, there seems to have been little ground for the extremely high hopes which had been entertained by the Committee which planned this division of the Investigation. The seven years which have intervened since that plan was drawn have witnessed great advance in the knowledge of the new-type test, with which at that time few, if any, of the social scientists had had much experience. They were, therefore, disposed to accept the great claims made for this new instrument of research not only by those who had devised the instrument, but also by some scholars in other subject-matter fields who had found it of great utility. Few, if any, of those who had developed the technique of the new-type test were scholars of the social sciences as well. They therefore had no basis for assuming that the new-type test could not be applied with as much success to the special field of the social sciences as elsewhere. Both social scientists and experts in educational measurement were, therefore, in a frame of mind which could only result in considerable disappointment as their joint work progressed—as social scientists learned more about the new-type test and the technician in measurement learned more about social science.

SUMMARY AND CONCLUSION

STATEMENT OF OBJECTIVES

Those technically interested in educational measurement desired from the social scientists a statement of ends, aims, and purposes of instruction in terms of objective knowledge, confident that once these objectives were so stated they could be definitely measured. Analysis of the materials and aims of social science instruction, however, indicated that no such simple solution of the problem of objectives was possible. The aim of social science was concerned not only with the acquisition of information or knowledge, but as much, or more, with the use of that knowledge in dealing with social matters to be faced by the pupil in adult life. The inclusion of such considerations of probability, possibility, and uncertainty introduced elements which seemed to defy complete objective statement. Furthermore, even the treatment of past knowledge, presumably completed and therefore apparently capable of objective statement, was found to involve questions of judgment. As Beard has pointed out¹, whether it was a matter of accepting past knowledge or of applying knowledge to present or impending social situations, certain ethical and esthetic considerations were nearly always involved. The individual's ideas of right and wrong, of expediency or in expediency, and his individual likes and dislikes played an important part both in his acceptance of knowledge of the past and in his use of knowledge for immediate needs. Knowledge of a purely objective nature alone did not determine conduct. It might condition conduct, and usually did so, but it was still possible for individuals possessed of the same knowledge to adopt quite different lines of conduct in a same social situation. Nearly every fact in social science was found to be the focus or nexus of a relationship which radiated out into the "seamless web" of world society. Nearly every idea was found to be part of a more or less extended pattern, which varied somewhat with each individual, and was more or less constantly changing. All this became increasingly apparent as work continued, and the actual trial of such tests as were constructed tended to confirm the above analysis. It was, there-

¹Beard, C. A.: *Nature of the Social Sciences*, vol. 7 in this series.

TESTS AND MEASUREMENTS

fore, impossible to state any major, and few minor, objectives of the social sciences in terms of objective or empirical knowledge. Such empirical data as social science employed represented but the beginning of social science, and the social scientist was primarily concerned with the subjective elements surrounding those data. At the risk of a play on words, it became clear that the most "objective" fact about social science was that it was so largely "subjective." The clear recognition of these conditioning elements proved a source of great disappointment to technicians in measurement, for thereby existing tests were rendered practically useless, and a modification of existing techniques was forced—a process necessarily slow and doomed to very limited achievement in the time allotted to the Investigation. The disappointment was shared by the social scientists.

GRADE PLACEMENT

Nearly every teacher of social science subjects feels that there is some natural order of progress in learning about society. It was hoped that the new-type test afforded the necessary instrument of research by which this order might be demonstrated. If so, the result promised would be of great aid to curriculum makers in this field, and equally so to teachers in determining their methods of instruction. "Grade placement" was the technical educational term applied, carrying obvious implications for both curriculum and method. Various efforts were made to discover the proper approach for the construction of tests.

Relative Difficulty of Courses in Social Science. The first approach attempted was to discover whether the several courses in social science presented inherent differences of difficulty. Examination of the numerous studies of curricula, published as well as unpublished, and of existing curricula in this and other countries failed to yield any decisive clues. The subjects generally employed in schools as the basis of later instruction in social science were found to be history, geography, elementary political science, economics, and sociology in various forms. Little, if any, uniformity was found in the arrangement of these courses in the schools. A course placed

SUMMARY AND CONCLUSION

last in the series in one school system was found very early in another. Oriental history, taught as a rule in the advanced years of college in this country, was taught to small children in China and Japan, and even in a few schools of this country. Community study, which by its immediacy might promise to serve as the first stepping stone, and actually was so used in many countries, appeared likewise as an advanced course in universities of this and other countries. Even geography, which seemed completed by the sixth or seventh grade in most schools of this country, was regularly taught in the secondary schools of Europe as well and in colleges both here and abroad. The more analytical social sciences, politics, economics, and sociology, which are usually more abstract as well as more analytical than history or geography, generally appeared late in the secondary schools, but important exceptions were found even in this arrangement. The only safe inference to be drawn from these studies was that no school course in social science was inherently more difficult than any other, as such, with the possible exception that the more analytical social sciences, when taught in a more abstract form, might be so regarded.

Relative Difficulty of Topics Within Courses. Having failed to find any decisive clues in the arrangement of courses, a more analytical study was made of the topics treated within courses. Analysis of courses of study had been a favorite theme for theses at most centers of educational research. Published and unpublished studies, the latter chiefly masters' theses, were examined. Most of these analyzed courses by topics with a nucleus of time, place, or homogeneous activity. These studies were supplemented by exploratory analyses of economic, political, sociological, and other elements.¹ Differences in degree of detail, the specificity of detail, the use of imaginative language and of emotional qualification were also canvassed. It became clear, however, that, as in the case of courses, topics likewise could not be classified by title alone as inherently of different order of difficulty. The Federal Reserve sys-

¹Dorothy Bovee, Dorothy Houston, Margaretha Jorgenson, Wilbur Murra, Renata Pecinovski, Lena Van Bibber and Olga Wold, all teachers of some experience, aided in this work. Miss Van Bibber's study was published in the *Second Yearbook of the National Council for Social Studies*, 1932.

TESTS AND MEASUREMENTS

tem of banking, reserved for intensive study in advanced courses of colleges and universities, was also treated in the elementary schools. If there were differences in difficulty as between topics, such difficulties lay in other aspects of the topic, not in the titles.

Relative Difficulty of Terms Used in Social Science. Disappointed both in the arrangement of courses and the topics within courses, attention was reluctantly turned to the study of the vocabulary of social science. Whether the reluctance was due to a widespread feeling that the study of words belonged to the linguistic sciences and not to the social sciences, or to a vague abhorrence of separate words as expressions of social ideas and occurrences, was never entirely clear. The words, however, did seem to present a definiteness and a precision which both the courses and topics had failed to yield. The existing word studies had been made for various purposes, linguistic, psychological, and educational. Some of those made out for spelling and reading contained ratings primarily on the basis of frequency of occurrence, though inferentially of some significance for relative difficulty also. There were also a few studies devoted to the special vocabulary of one or another of the social sciences. These included studies of concepts. The latter were found, upon analysis, to contain such complex matters as points of view, attitudes, and summary generalizations of large bodies of data concerning complicated problems. Much of this material offered the same difficulty found in the case of topics. Luella Pressey's study of the special vocabulary of history, Woody and Stephenson's list of terms in civics, and Eubank's list in sociology were more specific. These pioneer efforts were supplemented by the work of Edith Ware, Frances Baldwin, and Edgar Wesley, all on the staff of the Investigation. It was found possible to make a general distinction between terms designating such definite matters as persons, places, material objects, and events usually appearing as proper nouns and other terms which designated relationship. The latter appeared usually as common nouns or other parts of speech. Their peculiar value lay in their more universal applicability. The members of the staff directed their efforts to the com-

NATURAL ORDER OF LEARNING

pilation of such terms, as is indicated in the list contained in this volume.¹

There were disappointments in the study of these separate terms as there had been in the study of courses and topics. The precision of single words proved more illusory than real. Their definiteness tended to lie only in their spelling, and many of the terms were often used in a variety of meanings, and the same meaning was frequently expressed by a variety of terms. However, tests employing the listed terms were constructed and tried out, with some interesting results.

The Investigation was also aided in its study of grade placement by two scholars who were working independently on this problem. Professor Paul Lutz, of the Teachers College of Washington, D. C., had been at work upon the problem for several years both here and abroad. Professor D. S. Brainard, of the Minnesota State Teachers College at St. Cloud, was at work upon a study of grade placement in American History. Both generously aided the investigation. Lutz also made available his translation of the essay by Seignobos on this subject.

NATURAL ORDER OF LEARNING IN THE SOCIAL SCIENCES

Despite the numerous disappointments encountered in all this effort, the combination of analytical studies and tests yielded some results which may be of value. These findings are set forth tentatively, not through any lack of faith in their soundness, but because the matters with which they deal are both so subtle and so complex that others may desire a greater accumulation of evidence than can be here presented.

All matters of social concern have a location in place, and a setting in time, involving, of necessity, person and circumstance. Their relationship to other social occurrences, involves economic, political and cultural ideas. These ideas are expressions of social process in time, and yield yet further ideas applicable to future situations. The material elements are readily grasped, the other more slowly and in varying degree, according to the antecedent

¹See Appendix to this volume.

TESTS AND MEASUREMENTS

knowledge, interest, and competence of each individual. Thus, the recent London Economic Conference, whose potential significance could have been appreciated by only a very few, highly trained in the fields of international politics and finance, presented certain elements which even the most casual reader of the newspaper grasped. Even pupils in the elementary schools could name the conference, locate it, and mention the leading nations involved, and some of the chief actors.

The names of persons, places, and events are easily learned. Advance in knowledge consists in adding items of relationship. The degree of knowledge is measured, not by any particular item of relationship, but by the range of relationship or associations which the individual is able to connect with each name.

Terms expressing ideas may also be easily learned. Advance in knowledge of such terms consists in recognizing the shades of meaning which adhere to the terms in their various connections. A simple definition such as ordinary dictionaries afford usually represents the lowest common average meaning of the term. In actual use, the term means more or less than such a definition, and is seldom, if ever, completely satisfied by the definition. The degree of knowledge of any term is measured, not by recognition of a particular shade of meaning, but by the range of variation in meaning which the individual recognizes.

Knowledge of organized or closely related social phenomena proceeds usually from general to detailed knowledge, and from each separate phenomenon to its relations with other social phenomena. The possession of advanced knowledge is measured, not by the recognition of a particular detail or a particular relationship, but by the range of details and relationships.

Advance in social knowledge is a process of accumulation and integration varying greatly according to individual. Virtually every test tried out at different grade levels and even with groups homogeneous in physical or mental age, as the latter was determined by existing measures, revealed a fairly wide range of differences. One pupil was acquiring at one level knowledge which some of his fellows had already grasped three or four years earlier, and

NATURAL ORDER OF LEARNING

which others of his fellows were not to obtain for three or four more years. Since each grade level offered a different course or courses, the same knowledge apparently was acquired through different, or additional, media by the different pupils. This would seem to suggest that there is little basis for assigning any given course, topic, or term to any particular grade level.

The accumulation of social knowledge is, in large measure, dependent upon its integration with previous knowledge possessed by the pupil. The retention of new knowledge is probable chiefly in the degree to which it is related to existing knowledge. The pupil who is able to relate it immediately to his other social knowledge may require no further repetition, while those pupils who do not relate it may require many and varied repetitions before they retain the new knowledge. Each pupil may be regarded as building up a core of related knowledge. The complete integration of this information or knowledge to include all the elements listed in the second paragraph of this division is scarcely possible before the pupil has reached adult years. The process of integration on this more complete basis goes on through active life. In this sense of building up a core of completely integrated knowledge, school instruction in social science subjects can only serve as an introduction. The rounding out of that core, as well as its continued enlargement, is the work of adult years.

The relation of social knowledge to conduct is a more complex process than the acquisition of knowledge alone. Much of knowledge is acquired through the medium of words, whereas conduct involves additional emotional and motor elements, seldom translated into words. The individual's life is a microcosm of social activity. His education in social conduct involves both the recognition of this fact and the relation of verbal knowledge to his own experience and motor knowledge. Both are gradual processes and require sustained instructional effort in order that the pupil may identify his own experiences in their proper terms and thus link the more remote universal knowledge to these personal experiences. Only in this way can he be brought to make the fullest use of his growing knowledge in shaping his social conduct.

TESTS AND MEASUREMENTS

THE NEW-TYPE TEST IN THE SOCIAL SCIENCES

The Technique of the New-Type Test. The work of the Investigation brought the social scientist into close contact with the technique of the new-type test, as it related to his own subject. The apparent simplicity of the tests, with their short questions and shorter answers, had led social scientists to anticipate a very rapid production of tests for any of their purposes. They were much disappointed, therefore, by the fact that the most simple of the tests constructed for them required a year of time and great labor as well, while the very first of the tests projected, the Kelley-Trabue test, was not completed until the last year of the Investigation. Their disappointment was increased by the fact that the tests, when completed, touched so small a portion of the field of learning and seemed imperfect even for that area. The time and effort involved in the construction of the tests seemed out of proportion to the results obtained.

The form of the tests, designed apparently to limit answers to a single word and questions likewise to as few words as possible, also proved a source of disappointment to social scientists. Such tests seemed to imply a verbal precision for which social science offered little justification. Social ideas and social phenomena are usually capable of description and expression in a variety of ways. Furthermore, such minute sampling of social science knowledge clearly did not constitute a test of the student's comprehensive knowledge, or of his ability to develop sustained exposition of large ideas and to include the conditional elements which qualify any but the most simple of social situations. In other words, the extremely short answer form of the test seemed an artificial limitation which must confine such tests to the measurement of only the fragmentary beginnings of social science knowledge.

The statistical standardization of tests brought further disappointments to social scientists. The purpose of this standardization was to supply in the test an instrument applicable to all schools of the country at the particular grade level for which it had been designed. Earlier statistical analyses of the scores on the

NEW-TYPE TEST IN THE SOCIAL SCIENCES

trial of tests tended to eliminate those items which had proven very difficult or very easy. This process tended to defeat any possibility of testing the student's range of knowledge, and to confine measurement to mediocrity alone. It was not uncommon when the tests were applied to successive grades to find the more advanced students doing less well on many items than those in the earlier grades for whom the test had been standardized. The emphasis on the mediocre was penalizing those students whose knowledge had advanced to appreciation of shades of meaning, modifying conditions, greater degree of detail, and other qualifying considerations. Dissatisfaction at such an outcome is expressed in several of the articles in this volume.

These deficiencies in the technique of the new-type test as applied to the measurement of social science learning so marked in the early stages of this investigation have received attention from scholars in educational measurement. Anderson and Lindquist, working independently, have suggested important modifications in the technique of standardized tests applied to history and other social science subjects. The discussions by Kelley in the present volume show that the range of the tests can be greatly extended to satisfy many of the objections to narrow standardization. Further progress may be expected through the close co-operation of social scientists and scholars in educational measurement.

The New-Type Test as an Instrument of Instruction. The value of the new-type test as a supplementary device in school instruction, whether during the progress of the work or in the examinations at the close of instruction, seems established. Its deficiencies, however, are still so many that it can not be used as a substitute for all other tests at either of these stages.¹

The dangers arising from an exclusive use of the new-type test in instruction may be briefly summarized. Teachers seldom have either the time or the technical skill to prepare satisfactory tests. Hastily constructed tests are usually limited to those precise and separate material elements of instruction, the study of whose in-

¹For a more extended discussion of uses and limitations of new-type test in instruction, see pp. 95-113.

TESTS AND MEASUREMENTS

terrelationship, not the material elements themselves, constitutes social science. The student's tendency to determine the aims of instruction narrowly from the character of questions employed in tests and examinations would lead him thus to miss nearly all the aims of instruction in this field. One of the most striking losses resulting from excessive use of the new-type test is the student's failure to acquire practice in developing his ideas and knowledge in sustained, clear, and effective exposition. The other losses, though less striking, are just as real.

The chief uses of the new-type test in its variant forms as a supplementary device in instruction at present may be briefly enumerated. It is the most efficient device for detecting the student's possession of those separate material elements which, though not the end of instruction, are an essential preliminary to those ends comparable to the shoring which the engineer employs in shaping buildings made of concrete. No other testing device covers so great a range of information in so short a time, or can be graded so quickly and accurately. It may also be used to discover the student's knowledge of the simpler and limited relationship of this material. It may be used to some extent, also, in testing student's ability to apply ideas to new materials, and his possession of the skills involved in the subject. The more advanced and complex stages of these values, however, must as yet be discovered by other forms of test.

The New-Type Test as an Instrument of Educational Research. The new-type test was used in this Investigation primarily as an instrument of educational research in the field of the social science subjects. Its limitations, deficiencies, and failures as such have been sufficiently indicated and constitute an important finding of the Investigation. It remains now to consider the positive results of the effort.

The joint effort of scholars in social science and in educational measurement led to a more careful analysis of the aims and functions of social science instruction than has hitherto been available. Though the statement of objectives offered small comfort to those who had hoped that it might appear in the form of particularized,

NEW-TYPE TEST IN THE SOCIAL SCIENCES

material knowledge, that very fact directed the attention of workers in educational measurement to the less tangible, but equally real, values of the social science subjects. The recognition of the fact that the definite and material elements in social science were but means to ends and not ends in themselves was another important contribution of this work.

Clear recognition of the fact that social science was primarily concerned with the relationship of social phenomena and the acceptance of the implication of this fact gave a more definite direction to such pioneer efforts as those of Barr-Gifford, Woody-Stephenson and the Presseys. Their lists of terms, distinctive of the social sciences, could be pruned of too many purely material and specific elements and extended toward the inclusion of all terms expressive of relationship. Even though such a list of single terms, if completed, would be inadequate for the needs of social science, whose more advanced relationships can scarcely be stated in single terms or even in brief phrases, yet the use of such lists in tests may be regarded as a distinct advance.

Terms expressive of relationship are not limited to particular material elements or completed social phenomena, but are universal in their application to similar elements and situations present or impending. Knowledge of them, therefore, may be regarded as of permanent value to the possessor, a definite increment in his accumulation of social science. The Wesley, Wesley-Gold and Kelty-Moore tests constructed for this Investigation, and the Pressey tests, authorized as a technical check upon the former, show that considerable advance can be made in testing knowledge of relationship. The several ramifications of these tests, the most immediate and extensive contribution of the new-type test to educational research in social science subjects have been discussed at length earlier¹ and are therefore not repeated here.

Social scientists and scholars of educational measurement may not accept all the claims which the authors of these tests have made for them. Miss Kelty's insistence that pupils be drilled in these terms at early grade levels overlooks the fact that such

¹See pp. 95-113.

TESTS AND MEASUREMENTS

terms have a growing and widening meaning which is cumulative throughout the period of formal instruction, as well as later, and is built up, not by one course alone, but by all courses in social science, as well as by life's experiences outside the class-room. Her statement, however, adds needed emphasis to the fact that such terms are of vital importance in social science. The desired end may be gained even more effectively if, instead of drilling at *one* grade level, teachers at *all* grade levels give proper emphasis to the terms as they occur in their work, noting the different shade of meaning which the different situations involve.

The definite list of terms set down for separate grades by Mrs. Pressey carries an implication which her text corrects. The terms listed have cumulative meaning. At most, the significance of her particular grade lists may be stated as indicating that the larger fraction of pupils in the particular grades are aware of those words. The possibility that pupils at the fourth grade level might be made aware of almost any term, and that awareness of a term may imply nothing more than a recognition of that particular combination of letters is more clearly indicated in her text. Her tests, like the others, show the cumulative enlargement of knowledge of the relationships expressed by the term through successive grades.

The Clark tests and some of the Parker-Calkins tests, directed toward the measurement of skills, mark an advance in the testing of patterns of relationship. Whether such patterns are derived from the observation of recurrent complex social phenomena or represent logical schemes for the direction or control of social activity, they constitute some of the most important elements of social science instruction. Like single terms, few social patterns have sharply limited boundaries. In life itself they seldom recur in all details, yet they possess essentially common cores whose knowledge can be tested to a considerable extent. As these patterns represent more complex knowledge, the construction of adequate tests implies a greater demand upon the subject-matter specialists and suggests that the responsibility for further advance in the measurement of social patterns must rest primarily with the social scientist.

NEW-TYPE TEST IN THE SOCIAL SCIENCES

In the tests of attitudes and traits of personality, an effort was made to penetrate somewhat to those ethical and esthetic elements which condition nearly all social practice. Tests of attitude, long a preoccupation of psychologists, have been employed with some degree of success by social psychologists. The character of the attitude tests employed, however, and the limited extent of their application by this Investigation do not justify any definite conclusion as to their possible modification for use in the schools. The Jordan study of honesty and cheating in school situations confirms the impression of most social scientists that these qualities are not definite traits of personality, but, rather, reactions to immediate situations.

The general recognition of individual differences has served as a strong incentive to discover, if possible, the inherent qualities and traits of personality of which such individual differences are compounded. Though the social scientist is primarily interested in the social expression of personality, he would welcome the psychologist's discovery of any inherent and presumably fixed traits of personality as the basis of such social expression. It was with this hope that the Kelley-Trabue study was undertaken. Whether the traits selected are truly permanent traits of personality, and whether the results obtained are of value to social science, only further studies can demonstrate. The authors of the test are very modest in their claims for the results obtained. Their use of a free-association technique which permitted the subjects of the test the widest possible freedom of response must commend itself for its intellectual courage. Such procedure obviates many of the objections to the usual form of the new-type tests which limit the response of the subject within very narrow boundaries.

The use of the new-type test as an instrument of educational research in this Investigation has indicated that it can be made to yield additional data of value in conditioning educational policy and procedure. That data thus accumulated are not any more determinative of policy and procedure than data accumulated by other methods of research Beard has clearly indicated. The use of the new-type tests in social science subjects must be considered in

TESTS AND MEASUREMENTS

competition with other methods of research, and the prospects of returns from its use must be weighed against the great cost and effort involved.

Tentative Character of Present Conclusions. This volume is primarily a record of experience in the use of the new-type test in connection with social science instruction. The studies which it contains reveal the efforts of teachers and scholars in various subject fields of social science and of scholars in educational measurement. They reflect the background from which each of the authors has approached his task. So much progress has been made in the technique of testing, as well as in other fields of scholarship, since this work was begun that it is possible that many of those who have contributed to this volume would proceed quite differently were they to begin their work now. Those who have worked at the application of the new-type test to the problems of instruction in social science, either in this Investigation or *pari passu* with it, may be surprised that some of the ideas recorded in this volume were ever held by themselves. Yet the fact that they were so held suggests that others less experienced may also entertain them. A faithful record of our experience with tests in the social science field, however, revealing the story of hypotheses entertained and abandoned, conclusions reached and discarded, seemed in itself a contribution of value. If this record of our efforts enables future workers in this field not only to avoid the false steps here recorded, but also to utilize any advances indicated, it will be doubly useful.

APPENDIX I

A DIVERGENT OPINION AS TO THE FUNCTION OF TESTS AND TESTING

T. L. KELLEY

The use of tests in connection with the social studies has been widely discussed and has given rise to sharp differences of opinion. It seems desirable in this appendix that the background and resulting views of those who in general object to the use of tests in the social studies be stated in contradistinction to the background and views of those who in general advocate the use of objective tests in the social studies.

The writer has found it convenient to refer to the former as opponents and to the latter as proponents, and to use the letters "o" and "p" to designate these groups. The writer has ample reason to believe that the opponents as here characterized represent a very real and substantial group. He would classify himself as a proponent and has made bold to use the editorial "we" in stating the views of proponents for he has much reason to believe that the views so presented are widely held by experimentalists in the fields of psychology and the social studies.

It is hoped that antitheses drawn are so opposite that the views of opponents and proponents are fairly bracketed, but it is hardly to be expected that a comprehensive picture of any particular person's view will be entirely within the bracket thus formed. In fact, due to space limitations and attempts to simplify presentation, there are some very pertinent aspects of the view of the writer which are not touched upon (*e.g.*, the meaning and measurement of "reliability" as applied to trait and achievement measures). Though this is a report by a "proponent," it is not presented as an extreme view, but as the deliberate conviction of the writer.

We will then in short statements labelled (p) and (o) enumerate elements in the frame of reference of (p) the proponents, and in that of (o) the opponents.

TESTS AND MEASUREMENTS

(1-o) The opponents assert a close intimacy between teaching and testing regardless of the purpose of instruction.

(1-p) With this we agree but would emphasize these related uses of tests: (a) their use in the immediate facilitation of instruction of the pupils taking the test, (b) their use for the advancement of professional knowledge which will serve in the solution of various instructional problems, and (c) their use as instruments in any continuing, or fairly long time, guidance program. We would also emphasize that although every teaching or learning process implies a program of testing, it requires a separate focus of attention and frequently an additional and definite act to realize this program. To assume that the testing function is realized as soon as the teaching purpose is defined and content presented is education by faith. We specifically deprecate an educational program which does not involve some definite and adequate appraisal of pupil growth and accomplishment.

(2-o) The opponents manifest a general aversion to intelligence tests and all things for which they have been used, such as homogeneous grouping.

(2-p) In a number of fields including certain ones connected with the social studies, adequate knowledge as to the function of a general intelligence measure is as yet lacking, but we commend the use of the measure together with more specialized measures, for the purpose of ascertaining their limits of utility.

We admit to something of an aversion to intelligence tests, for the specific reason that they have been and are widely used in instances where purposes would be demonstrably better served by achievement tests of one sort or another. The intelligence measure is an omnibus affair that washes out detail. If the detail can be given with fair reliability by teachers' judgments or discriminating tests, we believe that it is always advantageous that it be given. We do not believe, e.g., that a homogeneous grouping or other adaptation to individual needs on the basis of any omnibus measure will meet ordinary classroom needs as well as a grouping or an adaptation which is more discriminatory of special abilities. The general intelligence test measure is the traditional measure,

APPENDIX I

and should be replaced in the social studies by new measures adapted to the particular problems of group guidance and instruction of the social studies. It should not, however, be replaced by judgments or new test measures greatly inferior to it in reliability.

(3-o) The opponents state that testers rate certain vocations as more *important* or *valuable* than others.

(3-p) This shows a misunderstanding of the significance of findings from various tests that there are substantial mean differences in accomplishment of members of various vocations. These mean differences do not determine the position of "an" engineer or of "a" skilled artisan. Neither do they imply anything as to the relative "importance" or "value" of the various vocations.

(4-o) The opponents show no awareness of the tests of reliability and validity of measuring instruments, either judgments of teachers or of test scores.

(4-p) We believe that such awareness is essential to any educator who is not content to work in the dark.

(5-o) The opponents show but little awareness of the distinction in function and limitation of different sorts of appraisals and tests, such as intelligence, achievement, character trait and attitude tests, etc.

(5-p) That teachers' appraisals of the achievement and capacities of pupils are characteristically woefully unreliable is accepted as a sad fact which cannot be remedied by any simple means and to which educators must adjust if a serious effort is made to adapt instruction and counselling to individual needs. That some of the so-called most important characteristics of people, namely their character traits, are still less reliably appraised either by judgments or by available tests, is a second sad fact to which one must adjust. It is futile to speak glibly of an education involving traits whose very existence in different amounts from pupil to pupil remains unknown. If certain trait appraisals have some, but only a little, credibility, it still remains relatively futile to build an elaborate educational structure upon them. The full extent and depth of our ignorance of character traits and their functioning is quite beyond the credibility of professors of the

TESTS AND MEASUREMENTS

social studies, and even of elementary and high school teachers who have not attempted to measure traits. Any structure built in ignorance of this is built upon quicksand, and accordingly efforts directed toward the better measurement of these traits are of prime importance, although such efforts be but partially successful.

(6-o) The opponents assert that objective tests can only measure the mechanical aspects of instruction and achievement.

(6-p) With this we unqualifiedly disagree. Any aspect of instruction subject to appraisal by teachers (and we affirm that all other aspects are pot shots in the dark and unworthy of being a part of recommended practice, but only of being objects of investigation) is ordinarily more accurately appraised by the consensus of several teachers or associates than by one; is more accurately appraised when judgments of these several are carefully combined into an average than when lumped together by "general impression;" and frequently is demonstrably more accurately appraised by tests of one sort or another than by the judgments of teachers and associates, even though two or three such judgments may be involved. These statements hold "on the average," for of course occasionally some single person's hunch will strike closer to the truth than any corresponding objective measurement, however refined, for complete universal accuracy is never to be expected.

In general the progressive lessening of the reliability of tests as the field of measurement moves from the mechanical to the spiritual aspects of a subject is granted, but it is probably not as great as the corresponding decrease in the reliability of teachers' judgments. This is a matter that can be experimentally determined in any specific case, provided that the trait in question is capable of being sensed by more than one person. The teacher claiming a unique and divine insight as to some characteristic of a pupil is clearly not subjecting himself to test, and equally clearly is making a claim that cannot be accepted by another as valid except it be taken on "authority." The first and greatest need in connection with modifying pupils with respect to some "higher"

APPENDIX I

aspect is knowledge of the certainty with which various techniques,—teachers' judgments, tests of various sorts, etc.,—indicate the existence of, or the amount of, these higher traits in pupils.

(7-o) The opponents assert that the use of objective tests diverts the attention from those aspects of education which are most important.

(7-p) With this we agree in the sense that test makers have frequently not pointed out the limited area covered by their tests, and that certain narrow-minded, easily satisfied teachers now exist, and may continue to exist under a new social order, who are prone to be content if the mechanical aspects of their subjects are mastered and if they have the supporting evidence given by tests that this is so. Such teachers will not be made efficient by withholding such instruments of precision as are available and as cover a part of their field of instruction. Such withholding would merely extend their field of inefficiency to include the more mechanical as well as the more subtle aspects of instruction. So far as it is possible the remedy for their narrow-mindedness lies in the criticisms of an intelligent supervisory staff.

For teachers with a vision beyond mechanics, trustworthy objective tests of the mechanical aspects will prove a boon. Were members of a class variable in such things as strength requisite to holding a piece of crayon, hearing sufficient to follow an ordinary conversation, eyesight required to read the common printed matter of the classroom, etc., the teacher simply asking a pupil to draw a map on the blackboard would find himself balked because pupils would fail, due to elementary or foundational conditioning factors. For the teacher to be unable to count upon strength to manipulate a piece of chalk, etc., would hinder the progress of any ordinary assignment in geography, and it would make impossible an assignment calling for undivided attention and a careful train of consecutive thinking. Only after all of these elementary conditioning things have been accounted for by homogeneous grouping of pupils,—by segregating or special handling of those lacking the requisite sufficient ability,—is it possible to devote undisturbed and concentrated thought to the intellectual problem which

is the matter of first importance. The expert teacher having adequate tests of vocabulary, or reading ability, etc., and having determined the mean status of the pupils of his grade, and the important individual differences existing with reference to knowledge of concepts, reading skill, and other mechanical factors, can now devote to the other issues of his subject an attention which is pointed and fruitful because conditioned by a knowledge of the foundation upon which to build. The objective measures have freed him from the bondage of devotion to the content they represent. Thus for the better teachers and the better methods of instruction the writers categorically disagree with the tenet that objective measures, even of mechanical phases, divert the attention from the more spiritual ends capable of realization through the subject.

The implication is here present that these tests of mechanical phases are not artifacts but measures of understandings and abilities which are needed in the existing world and will probably be needed in the world to which the pupil is being introduced. No defense is made of tests of which this is not true, and the determination of the fitness of the subject matter of the test is the joint responsibility of masters (1) of the subject in question, and (2) of the psychology of the levels of intelligence to which it is intended to present or teach the subject matter in question. While specifically affirming the necessity of subject matter specialists in connection with the grade selection made of test items, we equally affirm the necessity of first hand knowledge of the past achievement and present powers of children at the levels which it is proposed to teach and to test. We do not believe that the duties of the citizens-to-be are so simple in their demands upon understanding, so congruent with unmodified native tendencies, and so easy of execution that no problem of education exists except that of indoctrination and the measurement of loyalties. We look at specific indoctrination and specific loyalties as incidents in the flow of human life that will be found to be running its course long after these specific doctrines and attachments have been replaced by those of a still newer and more enlightened age. Meanwhile in the process of getting from the past to the future, we believe that

APPENDIX I

measurement of the loyalties considered highest at each step is an appropriate part of any tuition concerned with them. We would go further and advocate the measurement of collateral and antagonistic loyalties in order to enable an intelligent direction of whatever type of educative process is undertaken. We believe an attempt should be made at a definite evaluation of these loyalties, either by means of the judgments of competent people or by test measures which are demonstrably pertinent and of superior reliability.

The clause "demonstrably pertinent and of superior reliability" is the crux of this statement, for no test measure is axiomatically entitled to confidence, but only after its pertinence and reliability have been established or made highly probable by careful study. A granting of the entire argument of this report should not automatically lead to the belief in the efficacy of any single objective measure,—each must establish its field of usefulness.

Such belief should not rest upon the title of the test nor the name of its author but upon measurable agreement of test results with the judgments of teachers and other intelligent people,—the more the better,—acquainted with the persons tested. Though the test has a reliability indicated by a coefficient of .90, and teachers' judgments (for this trait in hand) of but .40, nevertheless the latter provide the standard whereby the validity of the test for the purpose in question is to be judged. Well established procedures for doing this are available.

Given, in an experimental situation, three measures for each pupil of a group: (a) a criterion measure of a trait X based upon the consensus of opinion of several competent people, (b) a single teacher's estimate of X, and (c) the score on an objective test purporting to measure X, and given the experimental finding that one of the last two measures tends to agree more closely with (a) than does the other. Then in a new situation wherein (b) and (c) only are available, it is the claim of test advocates that the measure which agreed the more closely with (a) is the more trustworthy measure of X. The ground for belief is thus simple and straightforward and has no mystical elements in it, and no prejudice in

TESTS AND MEASUREMENTS

it favoring the test measure as against the teacher's judgment measure.

(8-o) The opponents assert that individual differences are unimportant in connection with the broad purpose of a group concerned with fundamental curriculum revision in the social studies, which purpose should be primarily to create or prepare for a better social order.

(8-p) We unqualifiedly disagree. Directing the attention to individual differences in mathematical ability, musical ability, memory, self-control, affection, sociability, etc., there is evidence of disparities from individual to individual which would make such differences as those in height seem altogether trivial. If differences in mentality (either in general intelligence or specific abilities such as musical capacity, mathematical ability, memory, etc.) were reflected as differences in height, the world would be peopled with two foot dwarfs and twelve foot giants, with a heavy representation of all grades in between. This is the range of intellectual equipment with which we as educators are concerned, and to think of it as unimportant in connection with the so-called broad purposes of any group concerned with the fundamental problems of education is merely to characterize these purposes as pertinent to some other world than the one in which we live, or, with considerable plausibility, than the one in which our children and grandchildren will live.

The opponents conceive the purpose of such a group to be to create or prepare for a better social order. However, it is quite thinkable that a greater purpose would be to provide instruments for knowing the social structure, methods of appraisals of public servants, channels for the free flow of opinion and of information, and above all methods for experimentally verifying social hypotheses. Were this the purpose it would obviously change the entire effort of the undertaking which would then be to nurture rectitude of procedure and of thought in innumerable self-germinating centers of social reconstruction, rather than to be the initial germ center and the structural grandparent of all such endeavors.

It is claimed by the writer that a group committed to broad

APPENDIX I

purposes which do not include a study of individual differences is not qualified to pass judgment upon objective methods of measurement, for a correct appraisal of tests is to be expected only after a serious attempt to use them for their avowed ends has been made. The real criticism of a test is not to be made in terms of what the items "look like" on paper (suggestive though this may be) but in terms of the light the use of the test throws upon pupil development or behavior.

Objection to an approach dominated by the idea of experimental verification of conclusions is within the province of any deliberative body, which, however, must expect any of its own conclusions to be untrustworthy until they have been experimentally found tenable. The unremitting research and patience in awaiting outcomes of the scientist will alone serve to add the element of trustworthiness so necessary to pragmatic truth.¹ The words of John Dewey² well express the necessity for a fruiting in experience:

"In holding that the values which should determine the direction of education can be dug out of life-experience itself, we are denying by implication the position taken by some opposed types of philosophical theory. We affirm that genuine values and tenable ends and ideals are to be derived from what is found within the movement of experience. Hence we deny the views which assert that philosophy can derive them out of itself by excogitation, or that they can be derived from authority, human or supernatural, or from any transcendent source."

(9-o) The opponents assert that the forthcoming social order will be different from the present in important things of the spirit and that these cannot be measured by objective tests. Any reliance upon them is at best a concern with the trivialities of the new age.

¹This point is treated of in considerable detail in Chapter 7, *The Conference Method of Finding the Truth*, of *Scientific Method*, T. L. Kelley, 1932, Macmillan.

²Chapter 9, p. 294, *The Educational Frontier, Twenty-first Yearbook of the National Society of College Teachers of Education*, 1933. This yearbook, representing the conference product of a group of educational philosophers, is delightfully naive as coming from those whose slogan is integration, in its inconsistency from chapter to chapter. The present writer admits ability to quote from this same report statements of just the opposite vein from that of the above sentences of John Dewey.

TESTS AND MEASUREMENTS

(9-p) Granting a probable difference in social outcome as great even as a reversal of attitude toward private property, nevertheless, we do not grant a probable basic difference in the instincts, satisfactions, intelligences, skills, and types of social contact which will be ordinarily manifest in the new order. Further, we do not grant the impossibility of measuring by semi-objective means the new traits or accomplishments which will be deemed socially desirable. We do grant no present instruments for measuring these new things and very inadequate instruments for measuring most of the traits and accomplishments essential to happy living together in our present social order.

(10-o) As an estimate we would say that the opponents assert that the new elements in the mental life of the new social order bear a ratio to those in the present social order of about 9 to 1 in importance, if not in number and magnitude. By inference, no tests, however well adjusted at any given time, could ever be acceptable for a future time.

(10-p) We believe that tests can be changed in content and emphasis much more rapidly than the mental content of the mass of school or adult population can be changed, in this or in the new social order.

In the case of a growing thing, say a tree, the sap wood may bear a ratio to the heart wood of much less than one to nine, but still be of greater relative importance than nine to one in determining the nature of future growth. It, in short, is the only part that is growing. Does a parallel hold with reference to mental development? Presumably so with reference to changes in the views of adults wherein tradition, the heart wood, is massive. For example, thought and action in harmony with the capitalistic system may be second nature to the informed adult and not readily, or in the main, amenable to modification. To change such thought and action a tiny toe-hold via, say, a local cooperative store patronized but a few minutes a day may be more potent as determining growth than the mass of accepted ideas that daily control most of conduct. Here the new element in the individual's mental life, though bearing but a small ratio to the total, may bear a large

APPENDIX I

ratio to it as a determiner of growth. Should we apply the same view to school children? They are sprouts with negligible heart wood. They are undergoing a process of growing into the existent economic system in all its phases durable and transient. At the same time they are growing into innumerable other systems many of which, such as the spoken and written language of the country, tolerated conduct with reference to contacts with the stronger, the weaker, the brighter, the duller, the older, the younger, etc., call for an immeasurably greater thought and anguish than does the establishment of an attitude toward a church, a party, or a remote authority of any sort. The heavy assignments upon youth may be expected to continue into the new social order and the heavy duties of education may be expected to be the same in it as in the old order. The adult forgets the virginity of youth and that that which to him is prosaic and established is a vital phenomenon of life to the child. We therefore believe that to the infant, the child, the youth, and even to the late adolescent struggling to equip himself to render service that society will reward, there will be demands upon his mentality and his other developing powers which will be very similar to those at present. We do not subscribe to the idea that the demands in question are trivialities now nor in the generation to come.

(11-o) The opponents recognize no real need for proof of the attainability of any purpose as a prerequisite to the belief in and acceptance of that purpose.

(11-p) We assert that the first need of any objective concerned with the modification of school children is evidence as to the attainability of, or as to the conditions mediating attainability of, the changes in children postulated as congruent with the objective. We admit purposes lacking proof of attainability (more accurately tentative hypotheses) as appropriate to trivial issues, and those so urgent that time for study is not present, and as appropriate to investigational programs of all sorts, but not as appropriate to a program defended as well considered, desirable, and deeply affecting many people.

We grant that the attainability of remote purposes cannot be es-

TESTS AND MEASUREMENTS

established until the check-up at the remote time is made, but we affirm that until such check-up is made, the very reality of the purpose is questionable. A purpose that is uncertain is a purpose only in a dialectic sense, not a purpose in the world of sentient beings and present realities. That it is or is not attainable is unknown until the check-up has been made. The check-up is the other part of any serious undertaking, and a program that neglects it is not even half a program. A complete program involves purpose, check-up, and repercussion, of which the last is likely to be the most constructive element. We admit an aversion to remote and recondite purposes, and would at all times endeavor to phrase problems so as to be concerned with contiguous purposes and sequential progress from present realities and known resources to clearly stated mediate objectives. This is neither opportunism on the one hand nor fanaticism on the other.

(12-o) The experimental method is inapplicable as a means of investigating the broad purposes conceived to be important by serious educators, and the minor purposes are relatively trivial and not worth the thought of such educators in the present time of social stress.

(12-p) Such comprehensive purposes as consist of an effort to secure non-specific outcomes cannot be investigated by any means,—experimental, logical, or historical. It is only as they become defined in terms of specific ends or broken into aspects with definable consequences, that methods of investigation appear and become worthy of pursuit and of trust. Definable consequences are not verbalisms such as “to be a good citizen,” “to seek the social good,” etc., but rather acts of performance which can be observed, such as “going to the polls to vote,” “serving as a boy scout master,” “enumerating the constitutional duties of the governor of the state,” “participating in the sessions of the better government league,” etc. The labor involved is the only deterrent to investigating any purpose broad or narrow having specific outcomes in behavior, experimentally by means of carefully controlled conditions of sampling, of teaching, and of measurement. Usually, a synonym for broad purpose is indefinite purpose, and as long as the indefi-

APPENDIX I

niteness remains, we admit that the experimental method is inapplicable.

(13-o) The opponents ignore the uses and results of tests and other forms of appraisal in educational guidance and in vocational counselling.

(13-p) We consider educational and vocational counselling as a single function progressively becoming more important as scholastic differentiation takes place, and as higher and higher grade levels are reached. It is, we judge, the most intimate service to be rendered by the social studies. The inductions of the child into the political order as a citizen and into the economic order as a consumer have elements of similarity from pupil to pupil which are greatly in excess of those connected with his induction into the social order as a producer. The elements of similarity in the first two cases simplify instructional problems, but the importance to the child, and indirectly to society, of the third induction is so great that we consider an educational scheme that neglects this issue as derelict. As it enters in, it quite naturally becomes a part of the social studies program. To serve its needs an individual contact of rather a new order is called for between the teacher or counselor and pupil. We believe that this contact is of first importance and that objective tests, in so far as they are discriminatory and reliable, are of peculiar merit in its establishment upon a sound basis. Many tests that serve this purpose will also serve other educational purposes, but certain tests such as interest tests have a superior value here to that in connection with other educational practices.

(14-o) The opponents assert that "educators," psychologists and statisticians are trying to "run things" by means of tests of their own devising.

(14-p) One parent of this distrust may be incompetence of "educators," etc., but surely its other forebear is ignorance. Academic men in the field of the social studies in universities do distrust measuring instruments, the derivation of which they do not understand, and the reliability of which they cannot personally determine. However, as the only reason for this lack of ability

TESTS AND MEASUREMENTS

upon their part is their indisposition to study the techniques involved, their view is hardly entitled to great respect.

The willingness, the avidity, yes, the rather pitiable confidence, with which school men accept objectives given them by scholastic authorities, should absolve them from any desire to run things, except of course the machinery of the educational plant which they could not avoid running if they would. It is unfortunately true that "educators" poorly qualified in subject matter have rushed into print with subject matter tests. It is a matter of personal knowledge to the writer that in many instances where this has happened, it has followed a preceding serious but unsuccessful attempt upon the part of the educator to get the co-operation of academic authorities upon subject matter. The educator is frequently put in the position of being called upon to do something, and if university recluses giving courses upon the essentials of X, and how to teach X, refuse to concern themselves, either individually or in collaboration with public school men, with the measurement of X at elementary and high school grade levels, they should expect the educator to do something not to their liking.

The statistician is concerned with the construction of objective test only as with a minor chore.

The psychologist is brought into the picture, and rightly so, as an expert upon child capacities, just as is the subject matter specialist as an expert upon subject matter. The very nature of the problem calls for co-operative endeavor.

Generally speaking, there is a real difference between the attitudes of educator and social scientist. The former desires to co-operate, the latter to teach the value of a co-operative social order. However, we disapprove of any attempt to construct measuring instruments for different scholastic fields which do not involve the careful and critical assistance of competent specialists in these fields.

(15-o) The opponents fail to recognize that a multiplicity of semi-conflicting purposes and objectives should be encouraged; they also fail to recognize that a corresponding number of variously ~~directioned~~ tests is desirable.

APPENDIX I

(15-p) If the object is to help investigators ascertain consequences and purposes congruent therewith, or purposes and consequences congruent therewith, for it matters not which is conceived to be the starting point, it is apparent that many purposes will be postulated and many practices, including those actually or seemingly quite inharmonious, will be studied, and the significance and relationships which are imposed by the very structure of our psychological beings and by our human culture determined. The opponents think of efforts at social betterment and educational reconstruction as of the nature of an acid biting its way through to a pre-determined end, and accordingly one but half believed in, and fail to appreciate that efforts would probably serve a greater purpose if they were of the nature of a catalyzer, accelerating social growth and enabling reagents to combine into inherently harmonious patterns.

The writer has long maintained that for each of the disciplines represented in the social studies, different purposes, though each sponsored by reputable scholastic authorities, should be accepted as a basis for constructing objective tests, and that having these the status of school children be measured and the changes in status as correlated with changes in instructional techniques be experimentally determined, and that then and only then would one be in a position to formulate defensible and promising neo-educational doctrines of purpose, practice, and outcome. Present and future parties to curriculum reconstruction should feel a need for abundant and varied knowledge of individual and social facts and endeavor to supply this knowledge experimentally.

APPENDIX II

TERMS IN THE SOCIAL SCIENCES

This list is a sampling of terms expressing relationship. It is, therefore, neither a complete list of terms used in social science nor even of terms expressing relationship. Nor does it contain all the important terms expressing relationship. See Chap. 1.

Word	List Symbol ¹	Morton Weights ²	Wesley List ³	Wesley Tests, Editions No. ⁴					Wesley College			
				I	II	II½	III	IV	Unpub. ⁵	Pub. ⁶	Kelty ⁷	Pressey ⁸
abbey	1	-.227	x
abbot	1	-.029
abdicate	1, S	+.463	x	.	.	.	x	zz
abhor	1	+.168
ability	S
abolish	1	+.884	y
abolition	1	+.648	zz
aborigines	1	-.438	x
abridgment	x	.	.
abrogate	1	-.774	x	x
absentee	1	+.367
absentee landlord	1	+.367	x	.	.
absentee ownership	1	-.503
absolute ruler	x	.	.
absolution	2g	x
absolutism	1	-.714	x
absorption	1	+.215	x	.	.

¹The list symbols in this column indicate those lists in which a given term appears. These symbols have the following meanings: 1, The First American Historical Association List of Social Studies Terms; 2g, Miss Ware's Word List of Government and Law Terms; 2e, Miss Ware's Word List of Economic Terms; 2s, Miss Ware's Word List of Sociology Terms; E, Eubank's "The Concepts of Sociology"; and S, Stephenson's "The Special Vocabulary of Civics."

²This column shows the weightings given the terms in the First American Historical Association List of Social Terms.

³An "x" in this column indicates that the word also appeared in Dr. Wesley's List of Political Terms.

⁴"xs" in the five columns under this heading indicate that the term was used in those editions of Dr. Wesley's Test.

⁵An "x" in this column indicates words which occurred in the first version of the Wesley College Test in the Social Studies but were not included in the published test.

⁶"x" in this column indicates those terms used in the published form of the Wesley College Test in the Social Studies.

⁷An "x" in this column means that the term occurred in the original Kelty lists. A number shows that the term occurred in the original Kelty tests, the number indicating the grade level in which used. A "y" indicates that the term was used in the Kelty-Moore try-out forms. A "z" shows that the term is used in the Kelty-Moore final forms.

⁸"x" in this column marks words occurring in Dr. Pressey's original list of 1,444 technical words used in the social studies. This list was constructed from the following sources:

(1) An analysis of 23 textbooks in history and civics for use in grades 4-16.

(2) An analysis of 6 history texts widely used at the high school level, for which a frequency count was kept.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Katy	Pressey
abstract	1	- 222
academic freedom	1	-1 213
academy	1	x
accede	1	- .417	x
acceptance	1, 2e	+ 479	x
accessible	1	+ .536
acclaim	1	+ .125
accolade	1	-1 633
accommodation	2s, E	x	.	.
accommodation groups	2s, E
accountable	1	x	.	.
account, checking	2e
account, special interest	2e
accounting	1, 2e	+ .110	x	x
accredited	1	x
acculturation	2s, E	x	.	.
accumulated reserves	1	x	.	.	.
accumulation	1	+ .573	x	.	.
accusation	1, S	+ 705	x
achievement	1, 2s, E	+ .742
achievement, collective	2s
acquiesce	1	- 221
acquired characteristic	1	-1 263	x	x
acquisitiveness	1	x	x	.	x	.	.
acquisition	1	+ 120	x
acquit	1	.	x	x	x

- (3) Similar studies in history and civics: A. S. Barr and C. W. Gifford, "The Vocabulary of American History," *The Journal of Educational Research*, XX, 103-121, 1929; and A. W. Stephenson, "The Special Vocabulary of Civics," *The Journal of Educational Research*, XVIII, 297-304, 1928.

- (4) Frequency counts of special words occurring on the front pages of newspapers, in the editorials, and in many articles in magazines dealing with current events.

"y" marks words included in the list of 415 constructed from the original list of 1444, using the following criteria.

- (1) Frequency of appearance.
- (2) Combination of ratings made by 64 teachers of secondary school history, and 5 professors of college history, as to whether the words were "essential," "accessory," or "unimportant."
- (3) Ratings by 7 especially trained individuals (a college professor whose specialty is professionalized subject-matter in history, one college professor of history, an experienced and intelligent social worker, a mature high school principal who had taught history for several years, a public school specialist in educational research, a specialist in test construction, and the author of the tests) as to "sociological value" of each word,—that is, its use outside the history classroom.

Any words not meeting all three criteria were eliminated from this list of absolute essentials.

"m" indicates words included in try-out Forms A, B, C, D, but not in final forms A and B.

"Z" designates words included in final Forms A and B of the test, with the following qualifications:

"z1" marks words used as sample

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
act	2g, S	z
action (activity)												
corporate	E, 2s	x	.	.	.
actionable	x	.	.	.
activities	2s, S	x
adaptation	2s, E	x	x
adherent	1	-.206	x	x	x	x	.	x	.	.	.	x
adjourn	1, S	+.584	x	x	z
adjudge	1	-.613	x
adjust	1	+.584
adjustment	2s, E	x	.	.
administer	1, S	+.1731	x
administration	1, S, 2g	+.347	x	z
administrative												
department	(2g)
administrative functions	x	.	.
administrative law	2g
admiral	1	+.490	x	x
admiralty	2g	x
admission	x
admission of new state	2g
adolescence	1	-.391
adoption	1	+.446	x	x
adult	S
adultery	1	-.223
ad valorem	1	-.913	x	.	.
advance	x
advancement	x	x
advertiser	2e
advice and consent	x	.	.
advisor	x
advocate	1	+.047	x
aerial	x
aeronautics	x
affairs	S	x
affection	1	+.547	6	.
affectional	x	.	.
affidavit	x
affiliation	1	-.069	x
affirm	1	+.400
African	x	.	.
aftermath	x
age	x
agency	1, S	+.5840	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelcy	Pressy
agenda												x
aggrandizement	1	- 801	x
aggregate income	x	.	.	.
aggregate mass				x	.	.	.
aggregation	2s, E											z
aggression					x
aggressor					x	.	.
agitation	1	+ 531	x	x
agitator	1	+ 405	x
agrarian	1	- 922	x	x
agreement			x	x
agricultural labor	1	+ 432
agricultural revolution	1	- 619
agriculture	1, S	+ 722	z
aide-de-camp			x
airplane	1	+ 289	x
alderman	1, 2g, S	+ 010	x	x	.	x
alien	1, 2g, S	+ 416	x	x	.	.	z
alienate	1	- 279
alignment				x
allegiance	1, 2g, S	+ 169	x	x
allegation			x
alleviate	1	- 454	x
alliance	1, 2s, S	..	x	z
alliance regime			x	.	.
allies	S		z
allot	1	- 269
allotment			x	.	.
almoner			x	.	.
almsgiving	1, S	+ 216
almshouse	1, S	+ 215	x	x
altar	1	+ 663
alteration	1	+ 353
altering		x	.	.
altitude			x
altruism	1, 2s	- .556	x	x
amalgamation	2s, E		x	.	.
ambassador	1, 2g, S	+ 095	x	x	.	z
ambush			x
amendment	1, 2g, S	+ 332	x	x	z
America	S
American Confederation		
of Labor	2e
Americanization	1	+ .174	x	x	.	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
									Unpub.	Pub.	Kelty	Pressey
				I	II	II½	III	IV				
amicable	x
amity	1	-.887
ammunition	1	+.853	x	x
amnesty	1, S	-.946	x	x	.	x
amortization	2e
amphitheatre	1	-.023
amusement park	1	+.942
anarchism	1, 2g, 2e, S	-.714	x	x	z
anarchists	x	.	.
ancestor	1	-.795	x	x
ancestor worship	x
anchorage	x
ancient	1	+.647	z
animal husbandry	1	-.828
animate	1	-.211
animism	1	-1.450	x	x	.	x	.	.
animosity	1	-.128
annals	1	-.335	x
annexation	1, S	-.341	x	x	z
annihilate	1	-.180	x
annual	x
annul	1	-.102	x	x	.	x
anointment	1	-.787	x	x	x
anonymity	1	-1.144
antagonism	1, 2s, E	+.247
anthropology	x	.	.	.
anthropoid	x	x	.	x	.	.
anthropomorphism	2s	x
anticipation	E
anti-federalist	S
antipathy	1	-.453
anti (as prefix)	1	+.352
anti-slavery	z
anti-social	1	-.450	x	x	.	x	.	.
antiquity	1	+.199	x
anti-trust laws	1	-.152	x	x	.	x
apartment house	1	+.794
apology	1	+.673
apostle	1	-.089
appanage	1	-.152
appeal	2g, S	x	x	.	y
appeal to the country	x	.	.
appellate jurisdiction	1, 2g	-1.14	x	x	x
application	x	.	.	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
appoint	z
appointee	1, S	-.063	x
appointment	1, 2g, S	+.747	x	x	x
apportion	1, S	+.395	x
apportionment of representation	2g
appraiser	2g	x	x	x	x	.	x
appreciation	1	+.505
apprehension	1	+.058
apprentice	1	-.238
approach	E
approbation	2s
"appropriate legislation"	2g
appropriations	1, 2g, 2s, S	+.300	x	x	.	z
approval	2s
arbitrage	2e
arbitrary	1	-.011	x	zz
arbitration	1, 2g, S, 2e	+.122	x	x	.	z
archbishop	1	-.477	x
archduke	1	-.597	x
architecture	1	+.321	x
archives	1	-.792
archon	1	-1.34	x	x
area	S	x
arena	1	+.099
argument	1	+.616
aristocracy	1, 2g	+.222	x	x	.	.
aristocrat	x
armada	x
armament	1, 2g	+.179	x	z
armed neutrality	1	-.792	x
armistice	1, S	+.111	x	z
arms	x
army	S	y
armor	x
arrears	1	-.275
arraignment	x
arrest	S	x	x
arrogance	1	-.005
arsenal	1	+.205	x
arteries of traffic	1	+.057
art gallery	1	+.532
article	z
articles	S	x	x	x

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
articles of												
Confederation	2g, S	x
articulation	2s	x	.	.
artifacts	x	x	.	x	.	.
artillery	1	+ .579	x
artisan	1	- .269	x
art of living together	2s
art	2s, 1	- .085
ascetic	1	- .882
asceticism	2s
ascendency	E
aspersion	1	- 1.077
aspiration	1	+ .189
assailant	1	+ .288
assassination	1	+ .658	x	x
assault	x	.	.	x
assembly	1, 2g, S	+ .779	x	x	x	z
assent	1	+ .453	6	.
assertion	1	+ .505
assess	x	.	.	.
assessment	1, 2g, 2e, S	+ .237	x	z
assessor	1, 2g	+ .253	x
assets	1, 2e, S	- .005	x	x	.	x	.	x
assign	1	- .075
assignment	x	.	.	.
assimilate	1, 2s, E	- .012
assize	1	- 1.319	x	x	x	x	.	x
associate	1	+ .421	x
association	1, S, E	+ .616	x	x
assumpsit	2g
assumption	1	+ .132
assurance	1	+ .242
asylum	1, S	+ .427	x	x	.	.	.	x
atavism	x	x	.	x	.	.
atheist	1	- .413
at large, delegate or												
representative	S, 2g	x
atomization	E
atonement	1	- .755
atrociousness	x	.	.
atrocities	1	+ .116	x
attaché	S, 2e	x	x	x	.	x	x
attack	y
attainder	S, 2g	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
attendant	1	+ .558
attitude	1, 2s, E	+ .453	x	.	.
attorney	2g, S	x	.	x
attorney-general		x	.	x
auditor	1, 2g, 2e, S	+ .146	x	x	.	.	x	x	.	.	.	x
augment	1	- .178
auspices	1	- .063
Australian ballot	1, S, 2g	+ .644	x	x	.
authentic	x
authoritative statements	x	.	.
authorities	z
authority	1, 2s, E, S	+ .553	x
authorize	1	+ .453	x	z
autocracy	1, 2g	- .257	x	x	x	x	.	x
autocrat	1	- .115	x	x	.	x
autogeny	2s
autonomic behavior	x	.	.
autonomous	x	.	.
autonomy	1	- .919	x	x	x	x	x	x	.	.	.	x
auxiliary	1	- .136	x	.	.
aviation	1	+ .847	x
avocation	1	- .655
axiological	E
bachelor of arts	1	- .375
backward communities	x	.	.	.
backward country	x	.	.
backward regions	1	- .285
backwoodsman	1	+ .421
bail	1, 2g, S	+ .121	x	x	.	.	x	x
bailiff	1, S	- .865	x	x
balance of power	1, S	- .923	x	x	.	x
balance of trade	1, 2e	- 1.03
balance sheet	2e
balance, social	2s, E	z
ballot	2g, S
ballot box	2g
ban	1	- .293
band	x
bandit	x
banish	1	+ .474	x	x	x
bank	2e	y
bank, banker's	1, 2e	- 1.12
banker	x	.	.
bank, Federal Reserve	2e

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
bank reserve	1	-.660
bankrupt	z
bankruptcy	1, 2c	-1.479	x	.	.
bank solvency	1	-1.097
barbarian	1	+.042
barbarism	1, 2s	+.025	x
bargaining power	1	-.727
barge	1	+.016
barons	1	-.158	x
barracks	1	+.379	x
barrage	1	-.011
barricade	1	+.226
barriers	1	+.153	x
barrister	x
barter	1	+.274	z	x
base of operations	1	-.148
basilica	1	-1.187
basis of mutual aid	x	.	.	.
battalion	1	-.280	x
batteries	1	+.326	x
battle	y
battleship	1	+.515	x
bayonet	x
behavior	2s
behavior pattern	1, 2s, E	-1.732	x	.	.
beliefs	1	+.226	x	.	.	.
believers	1	+.226	6	.
bellicose	x	.	.
belligerent	1, 2g, S	+.053	x	x	.	zz
benefactor	1	+.311
bench	x
beneficiary	x	.	x
benefits	1	+.616
benefit of clergy	1	-1.138	x	x
benevolence	1	-1.60
benevolent despotism	1	-1.019	x
bench (King's)	1	-1.386	x
bequest	1	-1.437	x	.	.
besiege	1	+.642	x
bestow	1	+.279
betray	1	+.642
bey	1	-.887	x
bibles	1	+.500
bicameral legislature	1, S	-.914	x	x	.	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
bidder	1	+ 357
bids			x
bigamy	1	+ 190	x	.	.	.
big business	1	- 737
bigotry	1	- 324
bill	1, 2g, 2e, S	+ 516	x	x	z
bill board	1	+ 869
bill of attainder	1	-1 003	x
bill of exchange			x	.	.
Bill of Rights	1, 2g, S	- 697	x
bill drafting bureau	2g	
bimetallism	2e		x	.	x
biology			x	.	.
bi-partisan			x
birth certificate	1	+ 262
birth rates	1	- 149
bishop	1	+ 237	x
blacklist	1	- 133	x	.	.
blackmail	1	+ 005
blasphemy	1	- 133
bloc	1	- 851	x	x
blockade	1	+ 184	z
blockade runner	1	+ 173
block house	1	+ 237
blue laws	1	- 233	x
blue sky laws			x	.	.	.
bluster	1	+ 374
board	2g, S		.	.	.	x	x	y
board of directors	1	+ 137	.	.	.	x
boast	1	+ 721
body	S		x
bodyguard	1	+ 584	x
body politic	1	- 854	x	x	.	.
bogey	1	- 005	x
bogus			x
Bolshevist	1, 2e	- 369	x	x	.	.
bombardment	1	+ 406	zz
bombs			x
bonanza	1	- 714
bond	1, 2e, S	+ 047	.	.	.	x	x	.	.	x jr.h.s.	.	z
bondage	1	+ 316	x	x
bonding company			x	.	.	.
bondholder	1, 2e	+ 132	.	.	.	x	.	.	.	x	.	.
bond house		x	.	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
bond issue	1, S	-.190	x
bondsman	x
bonus	1, 2c	-.012	x	x	.	x
bonus system	1	-.582
bookkeeper	1	+.779
bookkeeping	x	.	.
boom	x
bootlegging	1	+.942	x	x	x
booty	1	+.189
border	x
borough	1, 2g	-.602	x	x	x	.	x	x
boss, political	1, 2g, S	+.811	x	.	.	.	x	x	.	x	.	x
boundaries	1, S	+.826	x	z
bounties	1	-.138	x
bounty, export	2c
bourgeoisie	1	-.893	x	x	x	.	x	x	.	.	.	x
boycott	1, S	-.328	x	.	x
Brahman priest	x
branding	1	+.305	5	.
branch	S	x
brassage	2c	x
breach	1	+.137	x	x
breach loading rifle	1	-.556
bread winner	1	+.663	x
breastwork	x
bribe (v.)	1, S	+.563	x	x	x jr.h.s.	.	.
bribery	1, S	+.506	x	z
bridge	1	+.837
brief	x
brig	x
brigade	1	-.191	x
brigadier	x
brigands	1	+.115	x
broadside	x
broker	1, 2c, S	-.201	x	.	x
bronze	1	+.231
brotherhood	1	+.495
browbeat	1	+.252
brutalities	1	+.595
buccaneer	1	-.719	x	x	.	x
budgetary reforms	x	.	.
budget	1, 2g, S	-.286	x	x	x	x	x	x	.	.	.	z
buffer state	1	-.719	x	x	.	x
building code	1	-.713

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelcy	Pressey
building and loan												
association	1	- .540
buildings	S	
bull (papal)	1	- 839	x	x
bull-baiting	1	- .956
bullion	1	- .596	x	.	.
bulwark	1	+ .258
buncombe	1	- .522	x	x
bureau	S, 2g	.	x	x
bureaucracy	1, 2g	- .997	x	x	.	x
bureau of standards										x	.	.
burgess	1, 2g, S	-1 009	x	x
burgher	1	- .956	y	x
business			z
business accounting	2e	
business cycle	1, 2e	-1 239
business depression	1	- 412
business efficiency	1	- 486
business fluctuation	2e
business forecasting	2e
business men			x	.	.
business risk	1	- .796
business unit	1	-1 239
buying methods	1	- .886
by-product	1	- .217
by-law	2g, S		x	x	x	x
cabal	1	-1 344	x	x	x	.	x
cabinet	1, 2g, S	- .177	x	z
cable	1	+ .426
cable transfer	2e	
cadet	1	+ .179
cahier	1	-1 06
calendar	2g	.	x	x	.	.	.
caliph			x	.	.
caliphate	1	-1 23	x	x
call money	2e
Calvinism	1	- 691
camouflage	1, 2g, S	+ 111
camp			x
campaign	1, 2g, S	+ 611	x	x	zz
campaign allowance	x	.	.
campaigner			x	.	.
campaign-fund	1, 2g	+ .431	x	x	.	.
campaign manager	x	.	.	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelcy	Pressey
campus	1	+.737
canal	5	x
canalization	2s,
cancel	1	+.721
cancellation of debt	1	-.017
cancelled check	1	+.179
candidate	1, 2g, S	+.647	x	x	x	z
cannibalism	1	+.516	6	.
cannon	4	x
cannonade	x
canon	1	-.691	x
canon law	1	-1.06	x	x	x	.	.
canonize	1	-.939
canton	1	-1.06	x	x	x	.	x
canvassing	x	.	.
capital (not econ.)	x	z
capital (econ.)	1, 2e, S	-.375	x	.	.	.	x	x	.	x	.	z
capital goods	x	.	.
capitalism	1, 2e	-.475
capitalist	1	-.074
capitalization	2e
capital punishment	1, 2g	+.058	x	x	x
capital ship	1	-.864
capitation tax	1	-1.25	x	x	x	x	x	x	.	x	.	.
capitol (pol.)	1, S	+.348	x	x
capitulary	1	-1.46
capitulate	1	-.556
captain	x
captain of industry	1	-.074
captivate	1	-.012
captive	x
capture	x
caravan	1	+.316	x
cardinal	1	-.347	x
career	1	+.574
cargo	x
caricature	1	+.063
cars and locomotives	x	.	.
carnage	1	-.175
carpet-bagger	x
cartoon	1	+.705	jr.h.s.	.
case	y
case method	1, S	-1.167
case study	x	x	.	x	.	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
case worker	1	-1 167
cashier			x	.	.	.
cash on hand			x	.	.	.
cast			x
castle	1	+ 558
careermen	2g
casualty		
catastrophe	1	+ 385
catechism	1	- 339	x	.
categories			x	x
cathedral	1	- 032	x
catholicism	1	+ 157	x
catholic orders	1	- 587
caucus	1, 2g, S	- 533	x	x	x	.	x	.	.	x	.	.
causation (social, natural)	2s, E		x	.	.	.
cause and effect			x	.	.	.
cavalier	1	- 612	x
cavalry			x
cede	1	- 195	x
celebration	1	+ 563
celibacy	1	- 692
censor			x
censorship	1	+ 442	x
censure	1	+ 385	x
census	1, 2g, S	+ 516	x	x
census enumerators			x	.	.
center of influence	1	- .633
center of population	1	- 211
centers			x
centralization	1, 2s	- 084	x
centralize	1	- .021	x
central government	1	+ .169	x
century		
ceremonialism	2s	
ceremony	1, 2s	+ 474
certificate	1	+ 574	x	x
certificate of ownership	2e	
certification of election	2g	
certify			x	x	x	.	x
cession	1	- .004	x	x	x	x	x	x	.	.	5	x
chagrin	1	+ 079
chain stores	1, 2e	+ 853
chain units			x	.	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
chairman	1, 2g	+.674	x	x	x	x
chamber
Chamber of Commerce	1, 2g, 2e	+.679	x
chamberlain	x
chancellor	1	-.677	x	x
chancery	2g	x
change of venue	x	.	.
changing	x	.	.
chaos	1	+.068
chapel	1	+.505	6	.
chaplain	2g
characteristic	x	.	.
charge (econ.)	1	+.108
charges (legal)	x
chargés d'affaires	1	-1.056	x	x
charity	1, 2s, S	+.616	x	4	x
chariot	1	+.600
charlatans	1	-1.018
chart	1	+.584
charter	1, 2g, S	+.122	x	x	.	x	.	z
chastity	1	+.126
château	x
chattel	1	-.570
chauffeur	1	+.753
chautauqua	1	+.194
chauvinism	1	-1.155
cheap money	1	-.977
check	2e, S	x	x	.	.	.
check, certified	2e	x
checking account	1	+.222	x
checkmate	1	-.426
checks and balances,
System of	1, 2g, S	-.876	x	x	.	x
chevalier	1	-.998
chief executive	1	+.185	x
chief justice	2g	x	x
chieftain	x	.	.
child guidance	1	-1.119
child labor	1, 2e	+.047
child labor laws	2e	x	x
child welfare	1	-.084	x	.
Chinese	x	.	.	.
chivalry	1	-.085
chores	1	+.511

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
christianize	1	+ 395
christen	1	- 015
chronicle	1	- 090
chronological age		
church	1, S	+ 795
church courts		
churchman	1	+ 237
circuit-court	2g, S		x	x	x	.	x	.	.	x	.	x
circuit-judge	1	- 459	x
circulating capital		
circulating medium	1, S	- 834
circumstances	1	+ 442
circumnavigation	1	- 054
citadel	1	- 174
city	2g, S	
city council	1	+ 437	x
city manager	1, 2g, S	+ 037	x
city planning	1, 2g, S	- 084	x	.	.
city social welfare		
activities	2g	
city state	1	- 528	x	x
citizen	1, 2g, S	+ 432	x
citizen army	1	- 089
citizenship	2g, S	
civic	1, S	- 348	x
civic center	S	
civil		
civil case	1	- 655	x
civil government	1	- 158	x
civilian	1	- 053	x	x	x	.	x
civil law	1, 2g	- 628	x
civil liberty	1	- 666	x
civil marriage	1	- 608	x
civil rights	1, 2g	- 714	x
civil service	1, 2g, S	+ 058	x
civil struggle	1	- 748
civilization	1, S	+ 179	x
civilized	S	
claimant	1	+ 278	x
claims	1	+ 579
claims adjuster	1	- 882
clan	1	- 258	x
clash		
class consciousness	1, 2s	- 770

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No					Wesley College			
									Unpub.	Pub.	Kelty	Pressey
				I	II	II½	III	IV				
classes									.	.	.	x
classic												x
classical			.			.		.		x		.
classified service	2g							.				.
class legislation	1	- 804	x
class struggle	1, 2s	- 661			
clause	1	+ 183	x								.	x
clearing house	2e					.	x			x	.	x
cleavage	1	- 265			
clemency						.	.	.				x
clergy	1	+ 373				.	.	.		jr h s		z
clericalism	1	- 902		x
clerk	S			
clerk of court	2g	
clerk of the House	2g	
client	1	+ 253	x
clientele	1	- 559	x	.	.
climate			x
clinic	S			
clique	1	- 012	x
closed season			x	5	.
closed shop	1, S	- .517
closure	2g	
clothing			x	.	.
cloture			x	.	.	.
club	S	
co-adaptation	2s, E	
coaling station	1	+ 195
coalition	2g	- 554	x	x	x	.	.	x
coast			z ¹
coast guard	2g		x
coasting vessel	1	- 248
code	1	+ 048	x	x
codify	1, 2g	- .485	x	x
coerce	1	- 236	x
cohesion, social	2s	
coinage	2e, 2s, 2g, S		x	.	.	zz
collaborate	1	- 368
collateral	1, S	- 481	x	.	x
colleague	1	+ 132	x
collections between												
banks	2e	
collective	1	+ 069	
collective bargaining	1, 2e	- .814			x	y	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
collective ownership	1	-.655
collectivism	2e, 2s
collector	S	x
college	x
colonel	x
colonial policy	1	-.516	x	x
colonies	1, 2g, S	+.416	x	x	.	x
"the color line"	1	-.344
co-maker	x	.	.
combat	x
combination	2s, S, E	x
commandeer	1	-.047
commander	y
commander-in-chief	1, 2g	+.616	x	x
commensurate	1	-.661
commerce	1, S	+.558	z
"commerce clause"	2g
commercial banking	1	-.677
commercial interests	1	-.339
commercial regulation	1	-.429
commercial rivalry	1	-.306
commercialization	2s
commissariat	1	-.859	z
commission government	1, 2g, S	.751	x	x
commission merchant	2e	x
commissioner	2g	x	x	.	.
commissioner
extraordinary	x	.	.
commissions (army)	1	-.206	x
commissions (group)	1	-.379
commit	x
committee	1, 2g, S	+.558	x	x	x	z	zz
Committee of the
Whole	1, 2g	-.912	x
commodities	1, S	+.384	z
common carrier	1, S	-.491	x	x	x
common council	2g
common defense	2g
commoner	x
common lands	1	-.593
common law	1, 2g, S	-.570	x	x	x	.	.	x
common people	x
common purpose	2s
commons	1	-.544	x

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelcy	Pressey
Commons (House of)	S
common stock
commonwealth	1, 2g, S	-.232	x	x	x	.	x	.	.	x	.	zz
commune	1	-1.088	x
communication	1, 2s, S, E	+.495	x	.	.	.	x	.	.	x	.	z
communion	1	+.052
Communion Supper	1	+.062
communism	1, 2s, 2g, E	-.527	x	x	.	.	x	x	x	.	.	z
communist	1	-.548	x	x	.	.
community	1, 2g, 2s, S, E	+.300	x	z
community center	1, S	+.205
community chest	1	+.105	x	.	.	.
community enterprise	x	.	.
community fund	x	.	.
commutation	x	.	.
commute	1, S	-.063
compact	1, 2g	+.247	x	x
companionship	2s
company	S	z
compatriots	1	-.185	x
compensation	1, 2s, S, E	+.311	x	.	.	x
compensation mechanism	E
competition	1, 2s, 2e, S	+.395	x	x	z
complainant	x
complex (cultural, social)	2s, E	x	.	.
composition (demotic, social)	2s	x	z
compromise	1, 2s, S, E	+.437	x	x	x
comptroller	1, 2g, S	-.861	x	x	x	x	.	x
compulsory	1	+.553
compulsory arbitration	1	-.525	x	.	.	.	x
compulsory education	1	+.426
concatenated	x	.	.
compurgation	1	-1.402
concede	zz
concentration	1	+.368
concern	x
concert	1	+.616
concerted volition	E
concession	1	+.142	x	.	x
conciliate	1	+.201	zz

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
conciliation	1, 2g	+.132	x	x	.	.
conclave	x
concord	1	-.012
concordat	1	-1.145	x	x	x	x
concurrent	S	x	x	x	.	x	x
concurrent jurisdiction	x	.	.
condemn	x
confederates	1	+.242	x	zz
confederation	1, 2s, 2g, S	+.158	x	x
Confederacy	1, 2g	+.163	x	z
conference	1, 2g, S	+.553	x	x	z
conference committee	2g
confession	1	+.484	x
confidence	1	+.642
confirm	1	+.379	x
confirmation	1	+.083	x	x	.	.
confiscate	1, 2e	+.258	x	.	.	.	x	zz
confiscatory	x	.	.
conflict (v.)	1, 2s, E	+.526
conflict, situation	E	x	x	x
conflicting	1	+.632
conform	1	+.358
congeries	1	-1.797
conglomeration	1	-.487
congregation	1, 2s, E	+.426	x
Congress	1, 2g, S	+.558	x	.	.	.	x	zz
Congressional Record	2g
congressional district	x	.	.
congressman	2g, S
conjoint	1	-1.181
conjugation	2s
connive	1	-.592
conquer	z	x
conquest	1, 2s	+ 553	x
consanguinity	x	x
conscience	1	+.332
consciousness, self	2s
consciousness, social	2s, E
consciousness of kind	2s, E
conscription	1, 2g, S	+.195	x	x	x
consecration	1	+.064
consensus	2s, E
conservation	1, 2g, S	+.068	x	.	z
conservative	1, S	+.011	x	x	.	z

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
conservatism	1, 2s, 2g	-.057	x
consignment	x
consol	2g
consistent mode of response	x	.	.
consolidated school	1	-.016
consolidation	1, 2e, 2s, S, E	+.005	x	x
conspiracies	1	+.321	x	x	x	.	x	x
conspirators	1	+.384	x
conspire	1	+.316
constable	1, 2g, S	+.358	x	x
constabulary	1, 2g, S	-.353	x
constancy	1	+.195	jr.h.s.	.	.
constituency	1, 2g	-.458	x	x	x	x	x	x
constituents	x	.	.
constitution	1, 2g	+.395	x	4	z
constitutional	x	.	.
constitutional amendments	x	.	.
constitutional basis	x	.	.
constitutional monarchy	1, S	-.449	x
constitutional revision	x	.	.
constitutional supremacy	x	.	.
construction (loose, strict)	1, 2g	+.463
consul	1, 2g	-.200	x	z
consul general	x	.	.
consul indorsement	x	.	.
consular service	1	-.405	x	x	x	.	x
consulate	1	-.274	x
consumer	x	.	z
consumption	1, 2e, S	+.489	x	.	.	.
consumption tax	1	-.769	x
contact	2s, E
contagion	2s
contagion, social	E
contagious disease	x	.	.	.
contemporary	1	+.225
contemporaries	x
contempt	1	+.616
contempt of court	1, 2g	-.580	x	x	.	.
contentious	x	.	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
contest	1	x
continent	1	+ .511	x
continental	1	+ .322	x
Continental Congress	S	x	x
contingent	1	- .702	x
continuity, social	2s, E	x	.	.
contraband	1, 2g	- .465	x	x	x	.	.	x
contract	2g, S	x	x	.	x
contraction	x	.	.
contract labor	1	- .628	x	x
contract, social	2s
contribution	1	+ .616	6	.
control, social	2s, E
control of industry by
workers	2e
controller of the
currency	2g
controversy	1	+ .442	x
convene	x	.
convenience goods	x	.	.
convent	1	+ .421
convention	1, 2g, S	+ .547	x	x	x	.	z
conventionality	2s
convergence	E
conversation	1	+ .542	4	.
conversion	1	+ .384	x
convey	1	+ .200
convict (v.)	1	+ .484	x	x	x
convict (n.)	1, 2g	+ .553	x	.	.	.	x	x	.	.	.	y
conviction (belief)	1, S	+ .253
convocation	1	- .665
convoke	1	- .723
convoy	1	- .539	x
cooperate	1	+ .516	x	x	.
cooperation	2e, 2s, S, E	x	.	.
cooperative marketing	1, 2e	- .454	x	.	.
cooperative societies	1	- .613
co-ordinate	1	- .205
co-ordinated governmental
agencies	x	.	.
co-ordination	x	.	.
co-partnership of capital
and labor	2e
cope	1	- .122

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
copperhead	1, 2g, S	-.122	x	.	.	.	x	x	.	x	.	x
copyright	1	-.713
cordon	1	+.178	x
coronation	1	+.036	x
coroner	1, 2g, S	+.099
corporal	1	+.110
corporal punishment	1
corporate action	2s, E
corporation	1, S	+.127	x	.	.	.	x	x	.	x	.	zz
corporation tax	1, 2g	-.638	x
corps	1	-.268
corpus delicti	x	.	.
correlary note	x	.	.
correlation	1	-1.007
correspondence	1	+.558	x
corrupt	1	+.558	x	x	.	z
corsairs	1	-.684
corvée	1	-1.360
cosmopolitanism	S, 1, 2s	-.560	x
cost belt	1	-1.333
cost of production	1, 2e	-.260
cotton-gin	1	+.532
council	1, 2g, S	+.547	x	.	.	.	x	x	.	.	.	z
councillor	x
counsel	2g	x	x
counselling	x	.	.
counsellors	1	+.368	x
counter-attack	1	+.237	x
counter-drive	x
counterfeit	1, 2g	+.663	x
counter-offensive	1	+.058
counterpoise	1	-.986
countersign	1	+.041
country	1, S	+.705	x	x	.	y
countryman	1	+.590
counts	1	+.295	x
county	2g, S	x	.	z
county seat	1	+.637	x	.	.	.	x	x
coup d'état	1, 2g	-.823	x	x
courage	1	+.837	4	.
courier	1	-.259
court	1, 2g	+.558	x	x	.	z
court-martial	1, 2g	-.099	x	x	x	.	x	.	.	x	.	x
court of appeal	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
court of records	1	..								x		
courtesy	1	+.721										
courtiers	1	+.016										x
covenant	1	-.374										x
covet	1	+.253										
covetous								x		
cowboy									5	
craft	1	+.316										
craftsman	1	+.068										
creative art	2s										
creative moment	E										
credentials	2g	x	x	x		x					x
credit	1, S	+.490										z
credit men								x		
creditor	1	+.306										
credit structure							x			
creed	1	+.178								x	x	z
creole	1	-.851										x
crew									4	x
crime	1, 2s, S	+.616	x	x	x							z
crime wave	1	+.183										
criminal	1, S	+.553	x						x			x
criminal law	1	-.096	x									
criminal offense	1	-.917	x									
criminology	1	-.402								x		
crisis	1, E	+.384										z
criterion	1	-.706										
critic	1	+.321									jr.h.s.	
criticism	1	+.505										
crop										x
cross fertilization of												
culture	2s										
cross questioning								x		
crowd	E										
crown										x
crown prince	1	+.384										
crucial situation	E										
cruisers	1	+.174										zz
crusades	1	-.137									jr.h.s.	z
crystalization	1	-.278										
cue	2s	+.131										
culprit	1	+.453										
cult	1	-.604							x			x
cultivation	1	+.574										

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
cultural lag	x	.	.	.
cultural stability	x	.	.
culture	1, 2s, E	+.390	x	.	.
culture mass	x	.	.	.
cumulative preferred
stocks	x	.	.
cupidity	1	-.535
curb	1	+.500	x
cure	1	+.748
curfew	1	+.179	x	5	.
curia	x	.	.	.
curial	1	-1.423
curiosity	1, 2s	+.689	4	.
currency	1, 2e, S	+.358	x	.	.	.	x	.	.	x	.	z
current	z
current expenses	x	.	.
curricula	1	-.165
curry	1	-.032
curtail	1	+.189
custody	1	+.184	x	x	x	.	x
custom (social)	1, 2s, E	+.443	x	.	.
customary codes	2s
customer	1	+.689	x	z	x
custom house	1	+.127
custom of the
Constitution	2g
customs (econ.)	1, 2g, S	-.126	z
customs duties	1	-.1995	x
cutthroat competition	1	-.282
-czar	x
cytology	x	.	.
Daimio	1	-1.781
dairy	1	+.674	4	.
damage	1	+.608
damage suit	1	+.352	x
damages, assess	2g
Darwinism	x	.	.
data	1	+.100
daunt	1	-.089	jr.h.s.	.	.
day nursery	1	+.153
deacon	1	+.179
deadlock	1	+.211	z
deal	x
dean	1	+.227

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
death penalty	1	+.779
death rate	1	+.194
debar	1	-.105
debase	1	+.011
debate	x
debauchery	1	-.138
debenture, export	2c
debt	S	zz
debt-funding	1, 2g	-1.014	x
debtor	x	.	.
debtor-class	1	-.466	x
decade	x
decadence	2s, E
decentralize	1	-.248	x
decarchy	1	-1.674	x
decision	1, 2g, S	+.6667	x	.	.	.	x	x	.	x	.	z
decisive	x
declaration	S	z
Declaration of												
Independence	1, 2g	+.850	x
declaration of war	1, 2g	+.733	x
decorate (medal)	1	+.567
decree	1	+.184	x	z
dedication	1	+.383	x	.
deductive process	x	.	.
deed	S	x
de facto	1, 2g	-.759	x
defalcation	x	.	x
default	1	-.029
defeat	x
defeatism	1	-1.136
defection	1	-.586
defectives	1, 2s, S	+.362
defendant	1, 2g, S	+.267	x	x	x	.	.	x
defense	1, 2g, s	+.711	x	x
defense mechanism	S
defensive	z
defiance	1	+.711	6	.
deficit	1	+.223
definition of situation	E
deflation	x
deforestation	x	.	.
degeneration	1, 2s, E	+.128
degradation	1, 2s	+.201

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
degree	1	+ .477
deist	1	- .970
de jure	2g	x
delegate	1, 2g, S	+ .633	x	x	z
delegated power	S	x	.	.	.
deliberate falsehood	x	.	.
deliberation	1	+ .372
delimitation	1	- .614
delinquent	1, S	+ .051	x	x
deliverer	1, 2s	+ .511
delusion	1	+ .211
demagogue	1, S	- .525	x	x	x	.	x	.	.	x	.	.
demagoguery	x	.	.
demands	1	+ .644
demand, alternative	2e
deme-enrollment	1	+1.647	x
dementia praecox	x	.	.
demobilize	1	- .011	x
democracy	1, 2s, 2g, S	+ .112	x	x	x	z	.
democrat	1	+ .439	x	z	.
democratic	x	.	.
Democratic Party	2g
demolish	1	+ .572
demonetization	1	-1.624	x
demonstration	1	+ .389
demoralize	1, 2s, E	+ .256
demote	1	- .007
denationalization	2s, E
denomination	1, E	+ .364	z
denounce	1	+ .511
density of population	1	- .303	x
department	1, 2g, S	+ .572	x	z
department store	1	+ .744
dependence	S	x
dependency	1, 2g, S	- .028	x	x	x	.	.
dependent (s)	1, 2s	+ .589
depopulation	1	- .107
deportation	1	+ .045	x	x
depose	1, S	+ .250	x	x
deposition	2g
depository	1, 2g	.000
deposits	1	+ .506	x	.	x
deposits as currency	2e
depreciation	2e	z

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kdty	Pressey
depreciate	1	- .139
depredations	1	+ .001
depression	1	+ .261	x jr.h.a.	x	.
deprive	1	+ .511	x
deputies	1, S	+ .294	x	x	.	.	x
deputized militia	x	.	.
deputy-governor	2g
dervish	1	- .831
descendant	1	+ .633	x	.	.	x	.	x
deseccration	x	.	.
desertion	1	+ .633	x
desire	E, 2s	x	.	.	.
desolate	1	+ .517	6	.
despoil	1	+ .107
despotism	1	+ .121	x	z
destroyer	x
destitution	1	+ .322	x
detachment	E
detective	x
detention	1	+ .317
deterioration	2s, E
determinism	1	- .831
deterrent	1	- .608
detestation	1	- .017
dethrone	1	+ .306	x	x
detour	1	+ .772
devastate	1	+ .322
development	1	+ .506
devolution	x	.	.
devotion	1	+ .711
devout	1	+ .517
diagnosis	1	+ .267
dialects	1	+ .182
dialogue	1	+ .339
diary	1	+ .517
dictator	1, 2g	+ .106	x	x	x
dictum	x	.	.
difference of potential	E
differential	x	.	.
differentiation	2s, E
diffusion	2s, E
dignitaries	1	- .011
diligence	1	+ .456
diminishing returns	1	-1.292	x	.	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
diminishing utility	1	-1.292	x	.	.
diocese	1	-.874
diplomacy	1, 2g, S	-.079	x	x	x	x	x	x	.	.	.	z
diplomat	1	-.338	x	x	.	.
diplomatic agents	1	-.338	x	x	.	.
diplomatic corps	1	-.338	x	x	.	.
diplomatic note	1	-.338	x	x	.	.
diplomatic relations	1	-.084	x
direct action	1	-.939
direct election	1, 2g	-.724	x
direct legislation	1	-.846	x	.	.	x	.	x
direct marketing	1	-.945
direct method	1	-.945
direct primary	1	-.840	x	x	x	.	.	.
direct sales	1	-1.072
direct taxes	1	-.761	x	x	.	.	x	x
director	1, 2e	+.289	x	x
director, interlocking	2e	+.289	x
directors, board of	2e	+.289
directory	1	+.179	x
dirigible	1	+.179	x
disabilities	1	-.011	x
disaffected	1	-.391	x	x
disarmament	1	+.379	x	z
disbanded	1	+.379	x
discipline	1	+.616	x	x
disconcerting	1	.000
discount	1	.000	x
discount of commercial	1	.000
paper	2e	.000
discourse	2s	.000
discover	1	.000	z
discredit	1	+.416
discrimination	1	+.263	x	.	.
disease	S	.000
disestablish	1	-.533	x
disfranchise	1	-.513	x	x	x	x	x	x	.	x	.	x
disinfectant	1	-.513	x	x	x	x	x	x	.	x	.	.
disintegrate	1	-.163
disintegration, social	2s, E	-.163
disloyalty	1	+.426	x	5	.
dismemberment	1	+.074	x
dismiss	1	+.074	x	x
disorganization, social	2s, E	+.074

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
dispatches	1	+ 311	x	x
dispensary	1, S	- 058	x
dispensation	1	.506	x	x	x
dispersion	2s
disposal	1	+ 253
disposition	1, S	+ 437
disputes	1, S	+ 568	x	x	4	x
disruption	1	+ 137	x
dissent	1	+ 121	x
dissolution	1, 2s	+ 016	x
dissolve	1	+ 447	x	x
distance, social	E
distinction	1	+ 437
distribute	1, E	+ 495
distribution, inequality												
of	2e
distribution (wealth)	1, 2e	+ 036
district	1, 2g, S	+ .558	x	y
district attorney	S	5	x
distrust	1	+ 568
disunion	1	+ 231	x
divergence	1	+ 179
diversification	1, 2s	- 264
diversity	1	+ 085
dividend	1	+ 264	x	.	x
divine right	1	- .212	x
division	2g	x
division of labor	1, 2s, S	- 201	x	.	.
division of spoils	1	- 195	x
divisions	1, 2g	+ 479	x
divorce	x
doctrine	1	+ 184	x	.	z
document	1	+ 253	x	x	.	z
dogma	1	- .628	x
dollars	1	+ .705
dollar diplomacy	1	- 729	x
domain	1	+ 015	x	x
domestic	1, S	+ 364	x	2
domestication	1, 2s	- 005
domestic manufactures	1	- .251
domestic policy	1	- .300	x
domestic regulations	1	- .490	x
domestic relations	1	+ 111	x	.	.	.	x
domestic system	1	- .744

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
									Unpub.	Pub.	Kelty	Pressey
				I	II	II½	III	IV				
domestic tranquillity	1	-.316
dominance	x	.	.
dominate	1, 2s, E	+.368
domineer	1	+.021
dominion	1	+.201	x	zz
dominions	1	+.126	x
donation	1	+.437	x	.
dormitories	1	+.705
dowager	1	-.436
dower	1	-.760
dowry	1	-.390
downfall	1	+.779	6	.
draft (army)	1, 2g, S	+.299	x	x	5	z
draft (money)	1, 2c	-.132	x
draftee	x	.	.	.
drafting	x	.	.
dragoon	1	-.486	x	x
draining	x	.	.
drawee	x	.	.	.
dreadnought	1	-.081	x
drug addict	1	+.068
drummer	x
dry	x
dual control	1	-.862	x
duchess	x
duchy	1	-.543	x	x	x	x	.	x
due process of law	1, 2g, S	-.846	x
duel	1	+.616	jr.h.s.	.
dues	1	+.558	x	.
dug-out	x
dukes	1	+.184	x	z	x
duma	1	-1.035	x
dumping	2c
dungeon	x	.	.
duplication	1	+.321
duties	1, S	+.689	x
duties (revenue)	1	+.132	x	z
dynamic	E
dynamic forces	x	.	.
dynamic immobility	x	.	.
dynamic society	x	.	.
dynasty	1	-.259	x	x	x	x	.	x	.	.	.	x
earl	1	-.205	x
Eastern Question	1	-1.009	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
eccentricity	1	- 474										
ecclesiastical	1	- 318										
ecological patterns												
ecology, human	E									x		
economic												
economic activity										x		
economical	1	+ 200								x		
economic conditions												
economic conflict	1	- .611								x		
economic expansion	1	- 713										
economic factors												
economic imperialism	1, 2e	- .972	x							x		
economic independence												
of women	2e											
economic												
interdependence	1	-1 008										
economic laws												
economic penetration	2e						x	x	x			
economic rights												
economics	1	- 085								x		
economic stabilization												x
economizing									x			
economy	1	+ .137								x		
edicts	1	- .126	x								x	
edit	1	+ .106										x
editorial	1	+ .500										
education	1, 2s, S	+ .501										
education, agricultural	2e									x	z	z
education, technical	2e											
educational												
qualifications	1	- .186										
educators	1	+ .501										
efficiency	1, S	+ .426										
efficient government												
eight-hour day	1, 2e	+ .443					x			x		
elastic clause	S		x	x								
elastic currency	1	- .992										
election	1, 2g, S	+ .632	x								z	z
election, biennial	2g		x	x	x	x	x	x				
electioneering	2g											
election returns	2g		x									
elective	S		x									
electoral college	1, 2g, S	- .470	x									
electoral district	1	- .407	x									

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
electorate	1, 2g, S	-.115	x
electors	2g, S	x
electrocute	x
eleemosynary institutions	x	.	.
elementary education	1	+.422
eligible	1, S	+.437	x	x	.
elite	1	-.549
emanate	1	-.532
emancipate	1, S	+.363	x	x	x	x	x	x	.	.	.	z
emancipation of women	2s
embargo	1, 2g	-.322	x	x	x	x	.	x	.	.	.	z
embassy	1, 2g, S	-.147	x	x
embezzle	1	-.131
embitter	1	+.184	jr.h.s.	.	.
embodiment	1	-.017
embryology	x	.	.
emergency	1	+.500
emigrant	1, S	+.243	x	.	.
emigration	1	+.243	z
eminent domain	1, 2e, 2g, S	-.819	x	x	x	.	x
emissaries	1	-.433	x	x
emoluments	x	.	.
emotional certainty	x	.	.	.
emotional similarity	x	.	.	.
empanel a jury	S, 2g
emperor	1	+.232	x	x	x
empire	1, S	+.179	x	5	z
employee	1	+.558	x	z
employer	1, S	+.558	x	.	y
employers association	1	-.192
employment	1, S	+.689	5	x
employment agency	1	+.421
employment bureau, government	2g, S
emporium	1	-1.555
empower	x
empress	4	x
emulation	1, 2s	-.327
enabling act	2g	x	x	x	.	.
enactment	S, 1	-.063	x	x	x	z
encirclement policy	1	-1.334	x
enclosure	1	-.364
encounter	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
encroachment	1	+ 022										x
encyclopedia	1	+ 590										.
endogamy		
endorse	1, S	+ 363	x	x	.	.
endowment	1	+ 084
ends, social	2s, E	
enemy		
energy, social	2s, E	
enforce	1, S	+ 527	x
enfranchise	1, 2g, S	- 269	x	x
engagement		
engineer	1, S	+ 706
Engel's law
engrossing
engrossment	2g		x	x
enjoin	1	- 576
enlightened	1	+ 206
enlightened despot	1	- 661	x	x
enlightenment	1, 2s	+ 048
enlist	1	+ 705	x
enlistment	2g	
enmity	1	+ 353
enroll	1, 2g	+ 578
ensign	1	- 406	x	x	x	x	.	x
enslave	1	+ 416
entail	1	- 1 023
entanglement
entente	1	- 776	x	x
entente cordiale	1	- 966	x
enterprise	1, S, 2e	+ 526
entertainment	1	+ 716
entrench	1	+ 137	x
entrepreneur	2e, 1	- 1.398
entrust	1	+ 311
enumerated article	1	- 598
enumerated powers	1	- 883	x
enunciation	1	- 131
environment	1, 2s, S	+ 311
envoy	1	- 052	x
epidemic	1, 2s	+ 432
epidemics, social	E	
episcopal	1	- 529
epoch
equality	1, 2s, S, 2g	+ 705	x

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
equalization	2s, E	x
equilibration	2s, E
equilibration of nations	x	.	.	.
equilibrium	2s, E
equipment	1	+.705
equity	2e, 2g, S	x
era	1	+.316	x	z
error	1	+.721
escadrille	x
escape	1	+.837
espionage	1	-.628	x	x
espouse	1	-.586
establishment	1	+.495	x
established church	1	-.581	x	x
estate	1	+.442	x	.	x
esteem	1	+.326
esthetic	1, S	-.718
estimation	1	+.326	jr.h.s.	.	.
estrangement	2s, E
ethical relativity	x	.	.	.
ethical standards	1	-.824
ethics	1	-.572	x	.	.
ethnology	1	-1.262	x	.	.	.
etiquette	1	+.258	x	.	.	x	.	.
eugenics	1, 2s	-.585	x	x	.	x	.	.
eulogize	1	-.333
European	x	.	.	.
evacuate	1	+.132	x	zz
evaluation of property	2e
evening school	1	+.590
event	z
eviction	1	-.106
evidence	1, 2g	+.390	x	x	.	.	x	x
evils	1	+.747
evolution	1	+.130	x	.	.
exaltation	1	+.084
examination	1	+.689	x	.
exarchate	1	-1.623	x
ex cathedra	1	-1.402
excesses	1	+.195
excessive capital outlay	x	.	.
exchange	1, S	+.526	x
exchange, foreign	2e
exchange, produce	2e

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
exchange, stock	2e											
exchequer	1	-.660										x
excise	1, 2g, S	-.629	x	x								x
exclusion	2g											x
exclusive	1	+.505									x	
excommunicate	1	-.475										x
execute	S											z
executive	1, 2g, S	+.300	x								x	z
executive act											
executor	1	-.181	x							x		
exempt	S											
exemption	1	+.264										x
exhibition of emotion									x		
exile	1	+.632	x							x		z
ex officio	1, 2g	-.586	x	x	x		x					
exogamy									x		
exorbitant	1	+.063										
expansion	1, 2s, E, S	+.390	x							x		z
expatriate	S		x	x						x		x
expedient	1	+.005										
expedition	1	+.616	x	x								z
expel	1	+.626									x	
expenditures	1, S	+.689										z
experimental research									x		
experiments	1	+.558									5	
expert	1	+.595										
exploit (v.)	1	-.163	x									z
exploitation	1, 2s, E	-.232	x	x						x		
exploitation of labor									x		
exploration	1	+.689	x									y
explosive											x
exportation	1, S	+.458										z
export duty	1	+.064										
exposition	1	+.263										
ex post facto	2g		x									
exposure	1	+.453									jr. h. s.	
express riders	1	-.359										
extension course	1, S	-.085										
extent of market	1	-.824										
exterminate	1	+.326										
external taxation	1	-.919	x									
extinction	1	+.263										
extortion	1	+.126										
extra-constitutional	2g											

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
extradition	1, 2g, S	-.566	x	x	x	.	x
extraction	1	+.258
extra-legal	1, 2g	-.840	x
extra-territoriality	1, S	-1.056	x	x	.	x
extravagance	1	+.632
extremists	1	+.247	x
extricate	1	+.079
extrovert	x	.	.
eye-witness	1	+.632	x	.	.	.
Fabian	1	-1.276	x
face-value	1	-.195
facilitate	1	-.052
facilities	x
faction	1	+.190	x	x
factor	x
factory	S	y
factory system	1, 2e	-.254
faculty	1	+.426
fallacious	1	-.279
family	S
family expenditures	x	.	.	.
family pattern	x	.	.	.
fanaticism	1	-.137
Far East	1	-.121	x
farm	1	+.837
farm bureau	1, S	-.177
farmer	x	.	x
farm loan	S
farm produce	1	+.620	4	.
farmstead	1	+.125	jr.h.s.	.	.
Facists	1	-.555	x	x
fashion	2s
fastness	1	-.110	x
fatalism	1	-.691
favorable balance of trade	1	-.635
favorable nation treaty	x	.	.
favoritism	1	+.316
fealty	1	-.813	x
fecundity	1	-.877
federal	1, S	-.028	x	z	.
Federals	zz
federal aid	1	-.133	x
federal courts	2g

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
federal government	1, 2g	+ .304	x
federal judge	x	.	.
Federal Reserve Bank	2e	x	.	.
federal reserve system	1	- .317
federal union	1	- .202	x
federalism	2g, S	x	.	.
federation	1, 2g	+ .042	x	x
fee	1, 2g, 2e, S	+ .526	x	x	x	x
feeble-mindedness	1	+ .320
felony	1, 2g, S	- .254	x	x	x	x	.	x
feminist	x
feminist movement	x	.	.
fence viewer	2g
fermentation, social	2s, E
fertility	1	+ .374	6	x
fertilizer	1	+ .501
fetter	1	+ .263
feud	1	+ .311	x
feudalism	1	- .507	x	x
feudal lord	1, S	- .523	x
feudal tenure	1	- .655	x
fiasco	1	- .612
fiat money	x	.	.
fictitious money	x	.	.	.
fidelity	1	+ .253
fiduciary	x
hief	1	- .628	x	x
field marshall	x
fighting machine	1	+ .199	x
figure head	1	+ .384
filibustering	1, 2g, S	- .645	x	x	x	.	x
filing system	1	- .035
finance	1, S	+ .437	x	.	z
finance company	x	.	.
financier	1, 2e	+ .437	x	.	x	.	.	.
financing	x	.	.	.
fine	2e, 2g, S	x
fire arms	1	+ .705
fire protection	2g
firm	x
first-class power	1	- .451	x	x
fiscal	1	- .535	x	x
fisheries	x
fixation	2s, E

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
fixed capital	x	.	.	.
flank	x
flag	x
flagship	x
fleet	x
flood	x
floor
floor leader	2g, S	.	x	x	.	.	.
flotilla	x
fluctuate	1	+ 016
fluid capital	1	-1 072
focus of attention	2s, E
foe	x
folk-moot	1	-1 423	x
folkways	1, 2s, E	- 919
follower	1	+ 637
foment	1	- 302
food	x	.	.
foodstuffs	1	+ 484	jr.h.s.	.	.
forage	x
foray	1	- 428	x
forecasting of demand	x	.	.
force	S
forced labor	1	- 069
forced loans	1	- .480
forces (military)	y
forces (social)	2s, E
foreclosure	1	- 259	x
foreign	1, S	+ 674	x	z
foreign affairs	1	+ 374	x
foreign born	1	+ 674	x
foreign exchange	1	- 254	x
foreign policy	1, 2g	000	x
foreign relations	2g
foreign soldier	x	.	.
foreigners	1, S	+ 616	x	.	.	.	x	x
foreman of the jury	1, 2g	+ 616	x	x	x
forerunner	1	+ .326
foresight	x	.	.
forester	S	x
forgery	1, 2g	+ .369	x
formalism	2s
formal organization	x	.	.
fortifications	1, S	+ .437	z

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Presscy
fortress	x
fortuity of human development	x	.	.	.
forum	1	-.064
forty-niners	x
foster	1	+.143
foster home	1	-.30
foundation	1	+.574
franc	x
franchise	1, 2g	-.075	x	x	.	x
franchise privilege	2g
franking privilege	x	x	x	.	.
fraternity	1	+.479
fraternize	1	-.068
fratricide	x	.	.
fraud	1	+.448	z
fraudulent	x	.	.	.
free-booters	1	-.339	x
free competition	1	-.390
free-holders	1	-.829	x
free-thinker	1	-.654	x
free-trade	1	-.269	x
freedom	1, 2s, 2g, S	+.437	x	x	4	z
freedom of contract	1	-.972	x	.	.
freedom of the press	1, 2g	-.012	x
freedom of religion	1, 2g	+.127	x
freedom of speech	1, 2g	+.058	x
free government	1	-.133	x
freeholders	x
free land	1	+.061
freedmen	1	-.187	x
freemen	1	-.080	x	x
free peasants	1	-.496	x
French Chamber of Deputies	x	.	.	.
friars	1	-.607
friction	1	+.316
frigate	x
frontier	1	+.442	x	z
frontiersman	1	+.442	x	x
fugitive	1	+.616	x	x	x	.	.	x
function	1, S	-.137	x
function, social	E

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
functional cluster												
pattern	x	.	.
functional integration	x	.	.
functionalization	x	.	.
functional middleman	x	.	.	.
functional pattern	x	.	.
functionary	1	-.517
fundamental	x	.	.
fundamentalist	1	-.607	x
fundamental law	1, 2g	-.972	x
funding	1	-.783	x	.	.
funds	1	+.495	jr.h.s.	x
fur trade	4	.
fusion	1, E, 2s	-.069
fusion of cultures	2s
futility	1	+.053
futures, dealing in	2c	x	.	.
gage	1	+.147
gag law	1	-.761	x	x	x	.	x
gag resolution	x
gainful occupation	1	+.247
gallantry	1	+.126
galleon	1	-.859
gallery	1	+.495
gallows	1	+.658	x	6	.
gambling	1	+.779
gangster	1	+.589	x	.	x
garison	1	+.432	x	.	.	.	x	x
gateway	1	+.553
gauntlet	1	-.063
gendarme	x
geneological continuity	x	.	.
geneological history	x	.	.
general	z ¹
generalissimo	1	-1.308	x	x
generalized concepts	x	.	.	.
general search warrant	1	-.725	x
general staff	1, 2g	-.306
general store	1	+.432
general strikes	1	-.317
general welfare	1	+.069	x	.	.	.	x	x
generations	1, S	+.253	x	.	x	.	.	x
genesis, social	2s, E
genetic psychology	x	.	.	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Presey
genetics (structures)	1, 2s	-1 018	x	.	.
genius	1	+ 057
Gentile	x
gentlemen's agreement	1, 2g	- 576	x
gentry	1	- 603	x
geography	x
gerrymandering	1, 2g, S	- 866	x	x	.	x
gift	1	- 705
gilt-edged	x	.	.
ghetto	1	- 633	x
gladiator	1	+ 174
globe	x
gold bullion	2e
gold certificate	2e	x	x	.	x	.	.
gold coast	1	- 687
gold, international flow	2e
of	2e
gold reserve	1	- .698
gold standard	1, S	- 692	x	.	x
good will	1	- .076
goods	1, S	+ 431	x	.	z
gospel	1	+ 300
govern	S	x	z
government	1, 2s, 2g, S	+ 558	x	.	.	.	x	x	.	x	.	.
governmental functions	x	.	.
government interference	1	- .116	x
government operation	1	- 070	x
government ownership	1	+ .205	x	x	.
government services	x	.	.
governor	1, 2g, S	+ 616	x	z
governor general	1	+ .553	x	x	.	.
gradation	2s, E
graduated	x	.	.
graft	1, S	+ .553	x	z	z
grandchildren	x	.	.
grandee	x
grandfather clause	x	x	x	.	x	x	.	x	.	.
grand jury	1	- .127	x	x
granger	1	- .434	x
grant	1, 2g, S	+ .505	x	x	.	zz
gratification	2s
great charter	S	.	x	x	.	.
"great masters"	1	- .444
"great powers"	1	- .244	x	z	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
great seal	2g											
greed	1	+ .553
greenback	1, 2e	- .144	x	.	.	x	.	zz
gregariousness	1, 2s	- 1.018	x	.	.
Gresham's Law	2e		x
grievances	1	+ .264
group	2s, E, S	
group consciousness	2s	
group conflict	1	- .672	x
guarantee	1	+ .374
guard			x
guardian	1	- .553	x	.	.	.	x	.	.	.	z	x
gubernatorial campaign	2g		x	x	.	.
guerrilla warfare	1	- .276	x	x	x
guidance	1	+ .405	x	x
guide book	1	+ .437
guilds	1	- .334
guild socialism	1	- 1.218
guillotine	x
guilty	x
gun boat	x
gun powder			4	.
habeas corpus	1, S	- .714	x	x	x	x	.	x	.	.	.	x
habitation	1	+ .132
habits	S	
habitual criminal	1	+ .058	x
hamlet	1	+ .132
hand-bill	1	+ .463
handicraft	1	+ .005	x
hangar	x
hanger-on	1	+ .079	x
harangue	1	- .200	x	x	x	x	x	x
harass	x
harbor	1	+ .658	x
hardships	1	+ .716
hard times	1	+ .558	z	.
harem	1	+ .063
harmony	1	+ .384
harrow	5	.
harry (v.)	1	- .616
harvest	x
haul (long, short)	S	
head money	1	- .846
head of family	1	+ .437	4	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelcy	Pressey
headquarters	1	+ 674	x
head tax	1	- 328	x
health	1	+ 690
hearing	2g
heathen	1	+ 357	x	x	.	x
heckling	x	.	.	.
hedging
heir-apparent	1	- 572	x	x	x	.	.
hemisphere	x
hemp	1	+ 189
henchmen	x	.	.
hereditary	1	+ 205	x
hereditary nobility	1	- .139	x
hereditary similarity	x	.	.
hereditary traits	x	.	.
heresy	1	- 218	x	.	x
heretic	1	- 177
heritage	1	+ 110
heritage, social	2s, E
hermit	1	+ 379
hero	1	+ .616
heterogeneity	x	.	.	.
heterogeneous	1	- .756	x	.	.
hetman	1	-1 696
hierarchy	1	- 976	x	x	.	x
high cost of living	1	+ 120	x	.
highness	x
high seas	1	.000	x	x	.
highways	S	x	x
Hindu	x	.	.	.
hinterland	1	- 704	y
hog-reeve	2g
holding company	1, 2c	- .935	x	x
holdings	1	- .202
homage	1	- .186	x	x	x
home	S
home-rule	1, 2g, S	- .269	x	x	x	.	.
homemade	1	- .800
homestead	1, S	+ .258	x
homicide	1	+ .064	x	.	.	.	x	.	.	x	.	x
homogeneous	1, 2s	- .545	x	.	.
honorariums	x	.	.
honors	1	+ .647
horde	1	+ .153

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelcy	Pressey
horizontal combinations	x	.	.
hospitable	1	+ .458	6	.
hospital	1	+ .837
host	1	+ .584	x	.
hostile	1	+ .379	z
hostile attitude	x	.	.
hotel	1	+ .942
hours, maximum	2c
House	x
House of Commons	1	+ .124	x	x
House of Lords	1	+ .057	x	x
House of												
Representatives	2g, S	z
household industries	1	- .096
housing conditions	1, S	- .059	x
hovel	1	+ .158
humane	1	+ .206	x	x	.	.	6	.
humanism	1	- .892
humanitarianism	1	- .329
humanity	1, 2s	+ .262
human nature	2s, E, S
humiliation	1	+ .389
hundred	1	+ .139
hunger strike	1	- .084
hustings	x	.	.
hut	1	+ .633
hybrid	1	- .426
hygiene	1	+ .267	x	x	.	.	x	.
hypocrite	1	+ .182
hypothesis	1	- .362
hypothetical case	x	.	.
iconoclast	1	- .921
idea-forces	2s
idealist	x	.	.
idealization	2s, E
ideals	1, 2s, S	+ .395
ideas	1	+ .512	5	.
identification	2s, E
identity of interests	x	.	.
idiot	1	+ .512	x	.	.	.	6	.
idler	1	+ .512
idol	1	+ .456
idolatrous	1	+ .312
ignominy	1	- .349

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
illegal	1, S	+ 456	x	x			x	x	.	x	x	x
illicit traffic	1	- 224	x	x	x	x	.	x	.	.	.	x
illiteracy	1, 2g, S	+ 249	x			
illiterates	S			x
illusion	2s		
imbecile	1	+ 278
imitation	1, 2s, E	+ 673	x	.	.
immemorial			x
immersion	1	+ 006
immigrant	S	
immigration	1, 2g, 2e, S+	456	.		.		x	x	.	x	.	x
immigration, restricted	2e		.		.		x
immorality	1	+ 389	x	.	.	.
immortality	1	+ 327
immunity	1, 2g	- 127
impair	1	+ 153	x
impanel			x
impartiality	1	+ 395	jr.h.s	.
impassable			x
impeachment	1, 2g, S	- 110	x	x	x		x	x	.	.	z	z
imperial	1	- 005	x	x	x	x	x	x
imperialism	1, 2g, 2s, S-	715	x	zz
imperial	1	- 442
impersonality	1	- 444
implied powers	1	- 708	x	x	.	.
import	1	+ 406	z
import duty	2g, 2e	
importation	1	+ 395	x	.	.
impose	1	+ 453
impost duty	1	- 555
impoverishment	1	+ 022
impregnable	1	- 170	x	x
impression	1	+ 458	x	.
impressment	1	- 475	x	x	x
imprison	1	+ 574	x	x
improvement	1, S	+ 632	x	x	x	.	.
impunity	1	- 287	x	x	x	.	x
inalienable			x
inaccessible	1	+ 395
inauguration	1, S	+ 522	x	x	.	.	x	z
inbreeding	x	.	.	x	.	.
incentive	1	+ 337	x	.
incite	1	+ 095
inclosure	1	+ 027

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
inclusive	1	+ .285	x
income	1, S	+ .558	x
income, net	2e	x
income tax	1, 2e, 2g	+ .201	x	x
incompetence	1	+ .269
incontrovertability
incorporation	1, 2g	- .175	x	x
incrimination
incumbencies
incumbent	1	- .228
incursion
indebtedness	1	+ .401	x
indemnity	1, 2e, 2g, S	- .185	x	x	.	.	x	x	.	x	.	2
indenture	1	- .809	x	x
independence	1, S	+ .322	x	x	2
independent
indeterminate sentence	1, 2g, S	- .693	x
index of prosperity	1	-1.224
index number	1, 2e	- .866
indictment	1, 2g, S	- .391	x	x	x	.	x	x	.	x	.	x
indifferentism	2s
indirect tax	1, 2g	- .466	x	x
individual	2s, S, E
individual examples
individual initiative
individualism	1, 2s, 2g, 2e	- .798
individualists
indorse	1	+ .148
indorsement	1	+ .079
indorser	1	+ .016
indulgence	1	+ .195
industrial	1, S	+ .390
industrial accidents
industrial center	1	+ .195
industrial conflict	1	- .324
industrial democracy	1, 2e	-1.077	x
industrial depression	1	- .272
industrial interests	1	- .266
industrialism	1, 2s	- .550
industrial revolution	1, S	- .319
industrial "satellites"	1	-1.287
industrial union	1	- .707
Industrial Workers of the World	2e

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
industry	1, 2e, S	+ .385
industry, infant	1, 2e	- .403	x	x
industry, key	2e
industry, localization of	2e
inefficient	1	+ .453
inelastic currency	1	-1 .098
inequality	1	+ .390	x	jr.h.s.	.
inequality of distri- bution	2e
inequality, social	2s
inevitableness of socialism	x	.	.	.
infallibility	1	- .376
infant mortality	1	- .121
infantry	x
inferiority	1	+ .390
infidel	1	- .011
infiltration	2s, E
inflation	1	- .698	x	.	z
inflict	1	+ .374	x
influence (v.)	1, 2g	+ .553	x	x	.	x
influx	1	+ .022
informant	1	+ .195
infraction	1	- .106
infringement	1	+ .037	x	x
in-group	E
inhabitants	z
inherent rights	1	- .750	x
inheritance	1, 2s	+ .501
inheritance laws	x	.	.
inheritance tax	1, 2g	- .079	x	x
iniquity	1	+ .085
initiative	1, 2g, S	+ .079	x	x	x	x	x	x	.	x	x	x
injunction	1, 2g, 2e, S	- .576	x	x	x	x	.	x	.	x	.	z
injustice	1, 2s	+ .437	x	x
inland waterways	1	- .005
inmate	1	+ .085	x	x	.	x	.	.
inn	1	+ .532
innate tendency	1	- .903
innocuous	1	- .929
innovation	1	+ .116
inoculate	1	- .052
inquest	x	x
inquiries	x	.	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
inquisition	1	-.660	x	x	x	x	x	x	.	x	.	x
inroads	1	+.137
insane asylum	1	+.664
insanity	1	+.501
inscription	1	+.253
insolvency	2e	x	x	.	.	.	x
insolvent	x	.	.
inspection	1, 2g, S	+.426	x	x	.
inspector	S
inspiration	1, S	+.258	jr.h.s.	.	.
instability	1	+.022	x
instability of labor	x	.	.
instability, social	2s
installation	x
instalment	1	+.316	6	.
instalment buying	x	.	.
instinct	1	-.207
institute	1	+.137	y
institution	1, 2s, E, S	-.005	x	x	.	.
institutional care	x	.	.
instructions	x
instrument	1	+.505	x	5	.
instrument of govern- ment	2g
insubordination	x
"insular cases"	2g
insular	x
insurance	1	+.369	x
insurance companies	x	.	.
insurance, liability	2e	x	x
insurance, life	2e	x	x
insurance, marine	2e	x	x
insurance policies	x	.	.
"insure domestic tranquility"	2g
insurgents	2g	x
insurrection	1	+.253	x	z
intangible	1	-.115	x	.	.
intangible property	x	.	.	.
integrate	1	-.861
integration	2s, E	x	.	.
integrity	1	+.064
intellectual	1	+.264
intellectual horizon	1	-.950

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
intellectualism	2s											
intelligence	S									x		
intelligence quotient	1	-1.051										
intelligentsia	1	- .918										
intemperance	1	+ .385										
intensive farming	1	- .153										
intention	S											
interaction	1, 2s, E	- .289										
intercommunication	E											
intercourse	1	+ .206										
interdependent	1	+ .022					x	x		x		
interdict	1	- .739							x			
interest	E, S									x		
interest rates	1, 2e	+ .190										
interests	1, 2s	+ .185										x
interference	1	+ .432	x								x	
interior										x
interlocking												
directorates	1	-1.062										
intermarry	1	+ .185										
intermediary										x
intermittent war								x		
internal										z
internal affairs	1, S	- .058	x									
internal improvements	1	- .132	x									
internal taxation	1	- .311	x									
internal waterways	1	+ .137										
international	1, S	+ .311	x									z
international adminis-												
trative union	2g										
international												
arbitration	1	- .154	x									
international												
commission	1	- .455	x									
international congress	1	- .137	x									
international												
cooperation	1	- .155	x									
international court	1	- .079	x									
international exchange										x		
international law	1, 2g	- .264								x		
international peace	1	+ .069	x									
international relations	1	+ .127	x									
international trade	1	+ .127								x		
internationalism	2s										

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
interoceanic	1	-.459
interpellation	1	-1.125	x	.	.
interpenetration	2s, E
interpretation	1	+.016	x	.	.	.	x	x	.	.	.	x
interregnum	1	-.882	x	x	x	.	x
interstate	S	z
interstate commerce	1	+.121
Interstate Commerce Commission	1, 2e, S	-.466	x
interstate trade	2g	x	x	x	.	.	.
interstimulation	2s, E
interurban	1	-.065
intervention	1	-.052	x	x	x	x	x	x	.	.	.	x
intestate	1	-.745
intimidation	1	+.258
intolerance	1	+.322	Y
intrastate	1, S	-1.176	x	x	x	x	x	x
intrastate trade	2g
intrigue	1	+.264	x	x
intrinsic value	x	.	.
introduction	1	+.632	5	.
introspection	x	.	.
introvert	x	.	.	.
intruder	1	+.411	jr.h.s.	.
invalid	1	+.637	zz
invalidate	x	.	.
invasion	1	+.553
invention	1, 2s, S	+.616	z
inventory	2e	x	x
investigation	1, 2g	+.632	x	x	x	x	x
investment	1, S	+.352	x	.	z
investor	1	+.522	x	.	.
investiture	1	-.883	x
invincibility	1	-.042	x	x	.	x
inviolable area	x	.	.
inviolability	1	-.537	x	x	.	.
invisible government	2g
ipso facto	1	-.772
irony	1	+.009
irreconcilable	x
irrigation	1, S	+.574	x	4	z
Islam	x	.	.
island	x
isolate	1	+.137

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelcy	Preserv
isolation	2s, E											
issues	1	+ 206										
issuance of money	1	- 206										
italic										x		
items of trade, invisible	2e											
items of trade, visible	2e											
itinerant	1	- .312										
Jacobinism	1	- .974	x									
jail												x
Japanese										x		
Jew												x
jeweler										x		
jingoisism	1	- .512	x	x								
jobber	1, 2e	- .199					x	x	x			
joint control	1	- .132	x									
joint occupation	1	- .121	x									
joint resolution	1	- .247	x									
joint session	1	- .212	x									
joint stock company	1	- .644										
joker										x		
journalism	1	+ .453									jr.h.s.	
journeyman	1	- .422										x
joust	1	- .547										
judge	1, 2g, S	+ .753	x	x			x			x		22¹
judgment	1, 2g, S	+ .663	x							x		x
judicial	S		x									
judicial decision	1	+ .252	x									
judicial interpretation	1	- .470	x									
judicial powers										x		
judicial prejudice										x		
judicial reform										x		
judicial review	2g									x		
judiciary	1, 2g, S	- .148	x									
junction	1	+ .384									6.	
junker	1	-1.039										x
junta	1	-1.138										
jurisdiction	1, 2g, S	- .252	x	x	x		x	x		x		2
jurisprudence	1, 2g	- .536	x									
jurist	1	- .080	x									x
juror												x
jury	2g, S			x								2
jury, trial (by)	1, 2g	+ .390	x	x							x	
jus sanguinis	2g											
jus soli	2g											

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelcy	Pressey
justice	1, 2g, 2s, S+	369	x									z
justice of the peace	1, 2g, S	+ 253	x								jr h.s.	
justice, associate	2g		.									
justice preventive	2g		.									
justiciable	1	-1 079	x	x	x		x				.	.
justification	1	+ 337	.								.	.
juvenile	1, S	- 374	.				x	x			.	.
juvenile court	1, S	+ 247	x								.	x
kaaba										x		
kaiser	1	+ 127	x								z	x
keynote			x									.
khedive	1	- 827	x									.
kidnap	1	+ 495	x			x					.	.
kin										x		.
kindred	1	+ 321										.
king											z	y
kingdom	1	+ 547	x								z	x
king's aid			.							x		.
kinship	2s		.									.
kinsman	1	+ 384	.									.
knight	1	+ 242	x								.	x
know-nothings			.									x
Koran	1	- 391	.									.
labor	1, S	+ 501	.								.	z
labor, child	2e
labor, conditions of	2e							x	.	.
labor disputes, arbitra-			.								.	.
- tion in	2e
labor federation	1	- .212	.								.	.
labor inspectors							x	.	.
labor problem	1	+ 057	.								.	.
labor rent	1	-1.048	.								.	.
labor savings	1	+ .311	.								.	.
labor turnover	2e
labor union	1, S	+ .258	.							x	.	.
labor, unskilled	2e
laboratory	1	+ .448
laborites	1	- .969	x							.	.	x
lackey	1	- .580
lady	1	+ .558	.							.	.	x
laissez-faire	1, 2e	- .838	.				x	x		x	.	.
land						x	.	.	.
landed aristocrats	1	- .528	.							x	.	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
									Unpub.	Pub.	Kelsey	Pressey
				I	II	III½	III	IV				
land-grant college	1	- 769										.
land-locked	1	- 026										.
landlord	1	+ 700										.
land, marginal	2c											.
landmark	1	+ 311									6	.
land ownership	1	+ 327									5	.
lands									x		.	x
landslide											.	x
land survey	1	+ 211									.	.
land tax	1	+ 006	x								.	.
lapse									x		.	.
larceny	2g		x	x	x		x		x		.	x
large-scale production	1	+ 005							x		.	.
latent social pattern									x		.	.
latent usefulness									x		.	.
latinized									x		.	.
latitude											.	x
law	1, 2g, E, S	+ 501	x	x							z	z
lawful											.	x
lawlessness	1	+ 474	x								.	.
law-making	S										.	.
law of supply and demand	1	- 459									.	.
lawsuit									x		x	.
lawyer	1	+ 827	x	x							.	x
laxness	1	+ 269	x								.	.
layman	1	- 064									.	x
leadership	1, 2s, S	+ 664									.	x
league	1, 2g	+ 301	x								.	z
League of Nations	1, 2g, S	+ 253	x								5	x
lease	1	+ 253						x	x		.	x
leave to print			x								.	.
lecture	1	+ 626									x	.
"Left"	1	- 718	x								.	.
legacy	1	- 074									.	.
legal	S		x								x	y
legal increment											x	.
legalism	2s										.	.
legality											x	.
legal personality	1	- 696	x								.	.
legal reform	1	- 455	x								.	.
legal tender	1, 2g	- 443	x	x							x	.
legal term											x	.
legal writ											x	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
legate	1	-.717	x
legation	1, 2g, S	-.275	x	x	x
legend	1	+.190
legion	1	+.116	x	x
legislation	1, 2g, S	+.458	x	y	z
legislative assembly	1	+.122	x	zz
legislative reference	2g
legislative responsibility	x	.	.	.
legislator	1, 2g	+.332	x	x	.	.
legislature	2g, S	x	x	x	.	z
legitimacy	1	+.069	x	x
leisure	1, 2s, S	+.332	6	.
leisure class	x
lettre de cachet	1	-1.497	x
letter of credit	2c
letters of marque and reprisal	2g	x
levy	1, S	-.117	x	x
liabilities	1, 2c, S	+.062	x	x	.	.	.	x
libel	1, 2g, S	-.408	x	x	.	x
libelous	x	.	.	.
liberal	1	+.258	x	x	.	x
liberalism	S
liberalization	2s
liberation	1, 2s, E	+.116	x	x
liberty	1, 2g, S	+.448	x	x	x	z
liberum veto	1	-1.581	x
license	1, 2g, 2c, S	+.258	x	.	.	.	x	x	.	.	z	x
lieutenant	1	+.474	x	x
lieutenant governor	2g, S
life organization	E
like-mindedness	2s, E
limitation	1	+.311	x	x	x	x
limitation of output	1, 2c	-.598
limited monarchy	1	-.418	x	x
lineage	1	-.189
linguistic	1	-.285
liquidate	1	-.613	x	x	.	x	.	x
literacy	1, 2g	+.005	x	.	.
literature	x	.	.
litigant	x	.	.	.
litigation	2g	x	x	x	.	.
litigious	x	.	.	.
little theatre	1	-.315

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
livelihood	1	+ .079										
live stock										x		
living wage	1	- 249								x		
loan	1, 2e	+ 490					x	x		x		x
loan, collateral	2e						x	x		x		x
lobby	2g, S									x		x
lobbying										x		
lobbyist	1	- 233	x							x		
local											4	2
local affairs	1, S	+ .385	x									
local custom	1	+ 310										
local government	1	+ 310	x									
local option	2g, S											
lock-out	1	+ 289										x
lock-step	1	+ 169										
locomotive												x
lodge	1	+ 501										
logic	1	- 295										
log-rolling	1, 2g, S	-7 03	x	x								
longitude												x
lookout	S											x
loom										2	
loose										x		
loose-construction	1	- .524	x									
loot	1	+ .285										x
lords	1	+ 373	x									x
loss of citizenship	2g											
lotteries	1	+ 274										
love	1	+ .558									4	
lower class	1	+ .264										
lower house	1	+ .145	x									
lower taxes										x		
loyalist	1, S	- 541	x									x
loyalty	1, 2g, 2s	+ .384	x								5	x
lucrative	1	- 238					x	x		x		
lump of labor theory	2e											
lump sum	1	+ .201										
lust	1	+ .125										
luxuries									x		
luxury tax	1	- .138	x									
lynching	1, S	+ .379	x				x	x				x
machine										x		
machinery	1, S	+ .505							x			x
machine gun											x

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
machine politics	1	-.453	x
made work	x	.	.
magazine	1	+.505
magisterial district	x	.	.
magistrate	1, 2g, S	+.122	x	x	.
magnanimity	1	-.675
magnate	1	-.385
Magna Charta	S	x
mail	1, S	+.837	4	x
mail order house	1	+.779	x	x	.
mainland	1	+.442	x
maintenance	1, 2e	+.379	x	x
majesty	1	+.352	x	x
major	x
major domo	x	.	.
majority	1, 2g, S	+.516	x	x	x	x	x	z
majority rule	x	.	.
maladjustment	1, 2s, E	-.196	x	x
maldistribution	1, 2s	-.755
maldistribution of labor	x	.	.
malfeasance	x	.	.
Malthusian	1	-1.413	x	x
maltreat	1	-.032
make work policy	2e
malingering	2e
mammal	x	.	.
management	1	+.448
manager	1, S	+.705	x	x
mandamus	1, S	-1.182	x	x	x	.	x
mandarin	1	-1.023	x
mandate	1, 2g	-.703	x	x	x	.	x
mandatory	1, 2g	-.707	x
maneuver	1	-.047	x	x
manhood suffrage	1	+.241	x
manifestation, social	E
manifesto	1, 2g	-.651	x	x	x
manifest pattern	x	.	.
maniken	x	.	.
manipulation	1	+.011
mannish	x	.
manor	1	-.407	x	x
manorial court	1	-.908	x
manorial system	1	-.908
man power	1	+.258	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
mansion	1	+ .264
manslaughter	2g	.	x	x	.	.
manual labor	1	+ .601
manufacture	1	+ .558	z
manuscript	1	+ .322
marauder	x
march	x
marginal land	2e
marginal price	2e
marginal production	1	-1.108
marginal utility	1, 2e	-1.318	x	.	.
marginal vendibility	1	-1.450
marginal wage	2e
margins	1	-.881	x
margrave	1	-1.234	x	x
marine	1, 2g	+ .327	z
mariner	x
maritime	2g
maritime law	1	-.651	x
mark	1	+ .184
market	1, S	+ .435	z
market channels	x	.	.
marketing	S	x	.	.
market performer	x	.	.	.
market price	1	-.107
marque	x
marquis	1	-.970	x
marriage	1	+ .374	x	x
marriage rate	1	-.503	x	x
marshal	1, 2g, S	+ .236	x	x
martial	x
martial law	1, 2g, S	-.148	x	x
mart	1	-.402
martyr	1	x
marxist	1	-.929
mass action	1	-.624
massacre	1	+ .416	z	zz
masses	x
mass meeting	1	+ .327	x
mass production	1	-.312
master	1	+ .495	x
masterpiece	1	+ .442
materialism	1, 2s	-.581
material civilization	1	-.613

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
material well-being	1	-.403
mathematics	1	+.264
matronymic	1	-1.366
matrimony	1	+.322
mature	x	.	.
maturity	1	+.253	x
maximum	1	+.384	z	.
mayor	1, 2g	+.479	x	z	x
means of												
communication	1	+.322
measure	1, S	+.443	x	y
mechanical	1, S	+.500	6	.
mechanism, social	2s, E
mediaeval	1	+.0005	z
mediation	1, 2g	-.122	x	x
mediator	x
mediocrity	1	+.016
medium of exchange	1	-.403
mêlée	1	-.755
meliorism	x	.	.
melting pot	1	-.382	x	x
members	S	x
memorandum	x
memoirs	x
menace	1	+.199
men-at-arms	1	-.079	x
Mendelian	1	-1.175	x	x
mendicant	1	-.754
menshevik	1	-1.387	x
mental age	x	.	.	.
mentally unstable	x	.	.	.
mercantile	1	-.149	x
mercantilism	1, 2e	-.866	x	.	.
mercenaries	1	-.481	x	x	x	6	x
merchandise	z
merchant	x
merchant guild	1	-.814
merchant marine	1, 2g, S	-.111
merge	1	+.169
merger	2e	x	x	.	.	.	x
meridian	x
merit	1	+.247

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
message	2g		x									x
message, presidential	2g											
messenger	1	+ 695										
mestizo	1	-1 256										
Methodism	1	- 259										
methodology	E											
metropolis	S											
metropolitan	1	- 337										x
Metternichism	1	-1 204	x									
middle-ages	1	+ 126									z	
middle-class	1	+ 188										x
middle-man	1	- 118								x		x
migration	1, 2s	+ 17										z
militarism	1, 2g, S	+ 121	x				x					
militaristic										x		
military												z
military campaign	1	+ 385	x	x								
military dictator	1	+ 264	x									
military discipline	1	+ 264	x								x	
military league	1	- 186	x									
military posts	1	+ 074	x									
military power										x		
military régime										x		
militia	1, 2g, S	+ 195	x									z
mind, social	2s, E											
mineral resources	1	+ 443										x
mine owners										x		
mines	S											x
minimum	1	+ 316										
minimum wage	1, S	- 254										
minister plenipotentiary										x		
minister	1, S	+ 501	x									z
ministry												x
minor crime										x		
minority	1, S	+ 327	x									y
minority leader	.									x		
minutemen										x		
minutes												x
mint	1, 2g	+ 385										zz
mintage	.									x		
mint ratio										x		
minutes												x
miscegenation									x			.
misanthropy									x			.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
misdemeanor	2g, S		x	x						x		
misogamy										x		
misogyny									x			
missionary												z
mission												x
mistrial												x
mob	1	+ 448										z
mob psychology	1	- 249										
mobility	2s, E											
mobility of labor	1	-1 014								x		
mobilization	1	- 184										z
mode	1	- 128									5	
model	1	+ 332									x	
moderation	1	+ 426										
moderator	2g		x	x								
modern	S											x
modern family										x		
modernist	1	- 577										
modification	1, 2s, E	+ 315								x		
Mohammedanism	1	- 127								x		
mollify	1	- 217										
monarch	1, S	+ 448	x	x								x
monarchist			x		x	x	x	x				
monarchy	1, 2g, S	+ 448	x									z
monastery	1	+ 321									z	z
monasticism	1	- 313										
monetary	1	- 132										
monetary unit	2e						x	x				
money	S											
money economy										x		
money, fiduciary	2e						x					
monk												x
monogamy	1	- 390										
monometallists	...									x		
monopoly	1, 2e, S	+ 131								x		z
monotheism	1	- 558										
Monroe Doctrine	1, 2g, S	- 064	x									
monument	1	+ 553										
moonshiners	...											x
morale	1	- 238	x							x		
morality	1, 2s	+ 327								x		
morals									x		
moral standard	1	- 047										x
moratorium	.. .									x		

APPENDIX II

Word	List Symbol	Morton Weights	Wealey List	Wealey Tests, Editions No.					Wealey College			
				I	II	III½	III	IV	Unpub.	Pub.	Keltry	Pressey
mores	1, 2, E	- 1 056								x		
moron	1	- 581				.	x
mortality	1	+ 005			
mortgage	1, S	+ 257	x			x	.	x
Moslem			x	.	.
Mosque	1	- 458						
mote	1	- 665					
mothers' pensions	S						
motion (parliamentary)	S		x	x				
motive	1	+ 448						
movement	S								.	.	.	y
M P.									.	.	.	x
muck-raking	1, 2g	- 597	x	x				
mulatto									.	.	.	x
multilateral treaty	1, 2g	- 1 140	x						.	x	.	x
multiplication	2s							
municipal									.	.	.	z
municipality	1, S	- 189	x						.	x	.	.
municipal ownership									.	x	.	.
munitions	1	- 100	x						.	.	.	zz
murder	2g								.	x	.	x
musketeers	1	- 005	x					
musketry									.	.	.	x
mutation	1	- 919						x	.	x	.	.
muster									.	.	.	x
mutiny	1	- 369	x	x	x				.	.	.	x
mutual aid	2s								.	x	.	.
mutual consciousness of likeness and difference	2s							
mutual exchange									.	x	.	.
mutually favorable terms									.	x	.	.
mythology	1, 2s	+ 080							.	.	.	6
nabob	1	- 919	x	x	x	x	x		.	x	.	.
nation	2g, 2s, S								.	x	.	z
national									.	.	.	x
national assembly	1	- 207	x					
national bank	1, 2e	+ 126						x	x	.	.	.
national bank notes	2e							
national church	1	- 528						
national debt	1	+ 158						
national defense	1	+ 269	x					
national domain	1, 2g	- 296					
national government									.	x	.	.
national guard	1, 2g, S	+ 122	x					

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
nationalism	1, 2s	— .339	x	x
nationality												z
nationalization	1, 2s	— .517	x	z
national parks	2g											.
nationals	2g		x									.
national suffrage												.
native-born	1	+ .322							x	.	.	.
native soldier									x	.	.	z
natural boundaries	1	— .075	x
natural desires	2s	
natural evolution			x	.	.	.
natural interests			x	.	.	.
natural resources	1	+ .327	z
naturalism			x	.	.	.
naturalization	1, 2g, S	+ .127	x	.	.	.	x	z
naval			4	zz
naval stores	1	— .194	x
naval supremacy	1	+ .069	x
navigation	1, S	+ .327	z
navy			z
Near East	1	— .470	x
necessaries	x	.	.	.
"necessary and proper"												
clause	2g	
necessity of life	1	+ .073
needs	S		x	.	.	.
negligence	2g, S	
negotiate	1, 2g	+ .258	x	x	.	.	z
negro			z
negro suffrage			x	.	.	.
neighborhood	S	
nepotism	1, 2g	— 1.287	x	x	x
net earnings	1	— .630	x	.
neutral country			x	.	.	.
neutrality	1, 2g	+ .174	x	z
neutralize	1	— .253
neutrals	2g	
neutral zones	x	.	.	.
newspaper	x
nihilism	1	— 1.051	x
nihilists	x	.	.	.
nobles	1	+ .325	x	x
nobility	1	+ .251
nomad	1	— .291	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelcy	Pressey
nominal	1	- 491										
nominate	S											
nomination	1, 2g, S	+ 206	x								x	
nominee	2g, S											x
non-competent	1	- 485										
nonconformist	1	- 676	x									
non-enforcible										x		
non-importation	1	- 296										
non-intercourse	1	- 244	x									x
non-intercourse (act)	2g											
nonjuring	1	- 914	x									
nonpartisan	1, S	+ 005	x	x								x
non-sectarian	1	- 212										
non-social	E											
nordic	1	- 719										
normal school	1	+ 253										
normalcy	1	- 469										x
Northwest Ordinance	2g											
notables	1	- 189	x	x								
notary	1, 2g	- 005	x	x	x	x		x				
notary's seal										x		
note	1, S	+ .437								x		x
note, Federal Reserve	2e											
nouveaux riches	1	- .559										
novelty goods										x		
novice	1	- .334										
nuisance	1, 2g	- 442										
null	1	- .048										x
null and void	S		x									
nullify	1, 2g	+ .078	x									
nun												x
oath	1, 2g	+ .437	x							x	x	x
obedience	2s											
obey	S											
obiter dictum	2g											
object, social	E											
objective	1	- 004										
objector	1	+ .322										
obligations	1, S	+ .385								x		
obnoxious	1	+ 184									x	
obscene												
observatory	1	+ .327										
observer	1	+ .442	x									
obsolescence	2s											

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
obstacle	1	+ .574
obstruct	1	+ .401
occupation	1, S	+ .716	x	z	z
occupation, blind alley	2e	x	x
occupation, seasonal	2e	x	x
odium	1	- .317
offender	1, S	+ .632	x	z	.
offense	S	x	x	x
offensive	1	+ .368	z
offerings	1	+ .626
office	1, 2g, S	+ .779	x	x	x
office seeker	1	+ .668	x
officers	S	x	z ¹
official	1, S	+ .632	x	x	z
official ban	x	.	.	.
officialdom	x	.	.
official ruling	x	.	.
old age insurance	1	- .455
old age pensions	2e
old fashioned	1	+ .569
Old Régime	1	- .966	x	x	x	.	x
oligarchy	1, 2g	- .823	x	x	.	x
omnipotence	1	- .370
one-price system	1	- .813
onset	1	- .089
onslaught	1	+ .0795
ontogeny	x	x
open diplomacy	1	- .723	x	x	x	.	.
"open door"	1, S	-1.056	x
open field system	1	-1.166
open shop	1	- .560
operating capital	x	.	.	.
operation	1	+ .437	x
operation costs	x	.	.
operatives	1	+ .067
opinion	1	+ .579	x	.	x
opium	1	+ .327
opponent	1, 2g	+ .626	x	x	x	z
opportunist	1	- .635	x	x
opportunity	1	+ .574
opposition (pol.)	2g, 2s, E, 1	- .482	x	x
oppression	1	+ .495	x	z
optimism	1	+ .189
option	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
orator	1	+ .689	x
ordain	1	x
ordeal	1	+ .200
order	S
orders (v.)	1	+ .500	x	.	.	x
order of the day	2g	- .608	x
order, social	2s, E
ordinance	1, 2g	- .081	x	x	x
ordination	1	- .482
ore	1	x
organic law	1	- .914	x	x	x	.	x
organism	1, E	+ .005
organization	1, 2s, 2g, S	+ 443	x	x	x
organized labor	1	+ 216
organized opposition	1	x	.	.
organizing life principle	2s, 2e, E
origin	1	- .054
original jurisdiction	1	- 1.135	x
originality	1	+ .253	x	x	.
origin of state	2s
origins, human	2s
origins of society	1	x	.	.
orphanage	1	+ .443
orthodox	1	- .195	x	.	.
ossification, social	2s, E	x	.	.
ostentation	1	- .112
oust	1	+ .090
outbreak	1	x
outcast	1	+ .316
outdoor relief	1	x	.	.
outflank	1	x
out-group	E
outlaw	1	+ .364	x	x
outlawry of war	1	- .439	x
outlay	1	+ .143
outlet	1	x
outpost	1	+ .211	jr.h.s.	x
output	1	+ .258	x
outrage	1	- .458
ourvote	1	- .047	x
overcapitalization	1, 2e	- .825	x
overdraft	1	- .645
overhead	1	x	.	.
overhead charges	1	- .756

TESTS AND MEASUREMENTS

Word	List Symbo	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
overinvestment	1	-.572
overland	1	+.316
overlord	1	-.192	x	x
overpopulation	1	-.082	x
overproduction	1, 2e	-.077
overpower	1	-.100
over-ride	1	+.099	x
over-run	1	+.258	6	.
over-seen	1	+.148
over-seer	1, 2g	+.379
overspeeding in industry	1	-.704
overt	x
overthrow	1	+.721	x
overture	1	-.048
overwhelming	1	+.442	x
ownership	1, S	+.558
pacifist	S	x
pacify	1	+.327	x
pact	1	-.127	x	y
pagan	1	-.0005
paganism	x	.	.
pageant	1	+.253	x	.
pages	1	+.558
paleoliths	x	.	.
palliative	1	-.728
pamphlet	x
pamphleteer	1, 2g	-.603
Pan Americanism	1, 2g	-.904	x	z
panel	1, 2g, S	-.128	x	x	.	x
panic	1	-.443	z
papacy	1	.000	x	z
paper blockade	1	-.898	x	x
paper money	1, 2g	+.114
par	1	-.445
parade	1	+.626	x
parasite	1	+.068
parcel post	2g	x
pardon	1, 2g, S	+.558	x	x	x	x	x
parental domination	x	.	.
parents	x	.	.
parish	1, 2g, S	-.053	x
parity	2g
park	1, 2g	+.689
parley	1	+.022	x	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Keltry	Pressey
parliament	2g, S
parliamentary bodies
parliamentary												
government	1, 2g	- 735	x
parliamentary												
procedure	1	- 783	x
parole	1, 2g	+ 079	x
partial	1	+ 269
participation	1, 2s, E	+ 327	x
participant observer	E
partisan	1, 2g, S	- 064	x	x	x	x	x	x
partition	1	+ 268	x
partnership	1, 2s, S	+ 558
party	2g, S
party affiliations		
party conventions		
party leader	S	
party machine	1	- 395	x
party organization		
party politics	1, 2g	- 847	x
party system, multiple	2g
party system, two	2g
party worker		
par value	2e
Pasha	1	-1.118	x
pass			x	x
passive resistance	1, 2g	- 846	x
passport	1, 2g, S	+ 201	x
patent	2g, S
patent law	1	- .217	x
paternalism	1, 2g, 2s	- 793	x	x	x	.	x	x
paternalistic venture
paternity	1	+ .131
pathfinder	1	+ .384
pathology, social	2s, E
patriarch	1	- .493	x	x
patrician	1	- 603	x	x
patrimony	1	- .750
patriotism	1, 2g, 2s, S	+ .373	x	x
patrol	1	+ .264	x
patronage	1, 2g, 2s	+ 009	x
patroon	1	-1.539	x	x	x	.	x	x
pattern, social	2s, E
patterns of control

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
pauperism	1, 2e, 2s, S	+.050	x	.	x
pawnshop	1	+.385
payee	x	.	.	.
payment	1	+.616	x	.	4	.
payroll	1	+.616
peace	2g, 2s, S	z
peace, breach of the	2g
peace conference	1	+.201	x
peace movement	1	+.069	x
peaceful penetration	1	-.940	x
peasant	1	+.188	z
pedagogue	1	-.291
peerage	1	+.392	x
peers	1	+.444	x	x
penal colony	1	-.410	x
penal law	1	-.527	x
penal system	1	-.555	x	x
penalty	1, 2g, 2e	+.374	x	x	x	.	x	.	.	x	x	x
penance	1	-.333
pending	x
penetration	1	+.206	x
penitentiary	1, 2g, S	+.501	x	x	x
peninsula	1	+.442	x
penology	x	.	.
pension	1, 2g	+.501	x	x
pension system	2e
peon	1	-.670
peoples	1, S	+.553	x	y
per-capita	x	.	.
perfidy	1	-.360
periodical	1	+.385
periods	1, S	+.495	x	z
perjury	1, 2g	+.090	x	x	.	.
perpetrate	1	+.195
perpetuation	1, 2s	+.084
persecute	1	+.443	z
person	E
personal	S	x	.	.
personal declaration	2e
personal disorganization	E
personal exemption	2e
personality	1, 2s	+.253
personality patterns	E
personality type	E

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelcy	Pressey
personalization	2s, E											
personal	1	- 940	x									
personal liberty	1	- 370	x									
personal property	1, S	- 117	.									
personnel	1	- 185										
persuasion	1	+ 389										
perversion	2s, E											
pervert	1	- 338										
petition	1, 2g, S	+ 506	x								x	z
Petition of Right	S		.									
petit jury	1	- 798	x						x			
petroleum			.									x
petty monarch	1	- 472	x									
petty state	1	- 419	x									
phalanx	1	- 697	x	x								
pharaoh	1	- 512	x	x								
phenomenon	1	- 049										
philanthropy	1	+ 122					x	x				
phratry	1	-1 618					x		x			
physical proximity									x			
physician	1	+ 616										
physics	1	+ 327	.									
physiocratism	1	-1 355								x		
picket	2e						x	x				x
picketing	1	- 286	.	.								
picture writing	1	- 364	.	.								
piece rate	1	- 755	.				x	x				
piety	1	+ 074									jr.h.a.	
pilgrim		x
pilgrimage	1	+ 327										
pillage	1	+ 067	x							x		x
pillory	1	- 323	x									x
pilot	1	+ 495
pioneer	1	+ 442	5	z
piracy	1, 2g	+ 385	x		zz
plagiarism	x		
pitched battle	1	+ .205	x		
plains			x
plaintiff	2g, s	x	x	x	.		.
plan	S		z
plank	2g		5	z
plantation	1	+ 522		z
platform	1, 2g, S	+ 558	x	x	.	.	x	x	.	.		.
plaudit	1	- 191

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
play	2s
playground	2g, S	x
playground director	S
plea	1, 2g	+.316	x	z	x
plebeian	1	-.408	x
plebiscite	1, 2g, S	-.866	x	x	.	x
pledge	1	+.379	5	x
plenipotentiary	1	-.887	x	x	.	x
plot	1	+.500	x
plunder	x
pluralistic behavior	E
plurality	1, 2g	-.493	x	.	x	.	x	x
plural voting	1	-1.009	x
plutocracy	1, 2g, S	-.803	x	x	x	.	x	.	.	x	.	x
pocket veto	1	-.819	x
point of order	1, 2g	-.603	x	x	.	.
point of view	1	+.448	jr.h.s.	.
police	1, 2g, S	+.616	x	y
police captain	2g
police court	1, 2g	+.437	x
police power	1, 2g, S	-.423	x
policy	1, 2g, S	+.437	x	z
political	x
political activity	x	.	.
political appointees	x	.	.
political institution	1	-.324	x
political aspirations	x	.	.
political district	x	.	.
political genius	x	.	.
political liberty	1	-.287	x
political machine	1	-.095	x
political parties	1, 2g	+.178	x	x	.	z
political reform	x	.	.
political rights	2g	x	.	.
political science	1	-.455	x	x	.	.
political situation	x	.	.
politicians	1, S	+.447	x	x	.	.
politics	1, 2g	+.315	x	z
poll tax	1, 2g	-.128	x
polls	1, 2g, S	+.206	x	x	z
polyandry	1	-1.066	x	.	.	.
polygamy	1	-.175	x	.	x	.	.	.
polyglot	1	-1.025
polytheism	1	-.898

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
pomp	1	+ .137	6	.
pontificate	1	- .946	x
pooling	1, 2c, S	- .507	x
pool room	1	+ .500
poor farm	1	+ .326
poor house	1	+ .326
"poor white"	5	.
pope	1	+ .258	x	x
populace	1	+ .111	x
popular	x
popularity	1	+ .563
popular sovereignty	1	- .829	x
popular suffrage	x	.	.
popular vote	1	+ .189	x
population	1, 2s, E, S	+ .326	x	2
Populism	2g
Populist Party	2g
port	1	+ .390	x
portage	x
port dues	1	- .228
portfolio	x
pose	1	+ .016	x	.	.
posse	S	x	.	x
posse comitatus	2g	x	.	.
possessions	1	+ .558	x	x
post	x
postage	1	+ .747
postal air mail	2g
postal money order	2g
postal savings	1	- .049
postal savings bank	2g
postal service	S
postal system	1, S	+ .385
poster	1	+ .453
posterity	1	- .059	x	.	x	.	.
postmaster	1, 2g, S	+ .695
postmaster general	S
post office	S
post office (rural free delivery)	2g
post roads	1, 2g	- .117
postpone action	x	.	.
potentate	1	- .459	x	x
potential customers	x	.	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
potential economic												
demand	x	.	.
poundkeepers	2g	x	.	.
poverty	1, 2e, S	+.506	x	x	.	x	.	x
power	1, S	+.448	x
powerful	1	+.448	x	4	.
powers	2
powers of Congress	2g
powers of president	2g
powers of states	2g
practices	1	+.385	x
pragmatic	1	-1.235	x
prairies	2
prayer	1	+.448
preacher	1	+.616
preamble	1	-.206	x	x
precede	1	+.258
precedent	1, 2g, 2s	-.122	x	x	.	2
precept	1	-.254
precinct	1, 2g, S	+.032	x	x	.	.
precipitates	1	-.006	x	x
precision	1	+.016
predecessor	1	+.274	x	x	x
predestination	1	-.782	x	.	.
predict	1	+.385
predisposition	1, 2s	-.423
predominate	1	-.011
pre-eminence	1	-.074	x
pfefect	x
prefecture	1	-1.061	x
preference, selective	2s
preferential tariff	1	-1.035
preferential voting	1	-1.072	x
preferred stock	x	.	.
prehistoric	1	-.053	x	x	.	.
prejudice	1, 2s	+.374	x	.	.
prelate	x
premier	1, 2g	-.326	x	x	2
premises	x	.	.
premium	1, 2e	+.327	x
preparedness	1, 2g	+.327	x	x
pre-primary	x	.	.	.
prerequisites	x	.	.
prerogative	1, 2g	-.692	x	x	x	.	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelcy	Pressey
presbyters	1	- 987										
pre-school	1	- 248										
prescribe	1	+ 201										
presentment	2g											
preserve	1, 2s	+ 469										
president	1, 2g, S	+ 617	x	x							x	z ¹
press	1	+ 469										x
pressure, social	2s, E											
pretence	1	+ 279										
pretension	1	+ 100										
prestige	1, 2s, E	+ 004										x
pretext	1	+ 158	x									
prevention	1, S	+ 406	x									
preventive medicine	1	- 381										
previous question	1	- 803	x									x
pre-war									4	
price	2e											
price fixing	1	- 703										
price fluctuations	1	- 697										
price of exchange	1	-1 055										
price, wholesale	2e							x	x			
prices	1	+ 305										
priesthood	1	+ 137									6	x
priest-king	1	- 686	x								6	
primary	1, 2g, S	- 117										z
primary education	1	+ 095										
primary election	S		x							x		
primary, non partisan	..									x		
primary, open									x		
primary, preferential									x		
primary schools	1	+ .148										
primate	1	- 951										
prime minister	1, 2g, S	- 195	x									y
primitive	1	+ 015										x
primitive peoples									x		
primogeniture	1	-1 035								x		
prince	1	+ 448	x									x
princess	1	+ 448	x								z	zz ¹
principal	1, S	+ .501	x								z	
principality	1	- 318	x									x
principles	1	+ .385	x									
printing	1	+ 690										
printing-press	1	+ .690										
prior	1	+ .079										

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wealey List	Wealey Tests, Editions No					Wealey College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
priority										x		
prison	1, 2g, S	+ 706	x									x
prisoner												x
privacy	1	+ 443										
private	1, S	+ 506										y
privateer	1	- 191	x	x	x		x					x
private expenditures									x			
private ownership	1	+ 146	x									
private property	1	+ 277	x								z	
private rights	1	- 292	x									
private war	1	- 428	x									
privilege	1, 2g, 2s, S	+ 316	x								x	
privileged classes	1	- 270	x									x
probate	S		x									x
probate courts										x		
probation	1	- 054	x							x		x
probation officer	1	- 065	x									
problem	1, S	+ 506									4	
procedure	1, 2g	+ 258	x									
proceedings										x		
proceeds	1	+ 122										
process, cultural	2s, E									x		
processes of production	1	- 512										
procession	1	+ 500										
proclamation	1, 2g, S	+ 447	x	x	x	x	x	x			x	z
proctor	1	- 259	x									
procurator	1	-1 250	x	x								
prodigy	1	- 248										
product, social	2s, E											
produce exchange	2e											
produce	1, S	+ 506										
production	1, S	+ 506							x			z
productivity of labor	1	- 841										
profession	1	+ 316										
professional class	1	+ 182										
professionalization	2s, E											
professionalization of												
industry	1	-1 113										
professor	1	+ 501										
proficiency	1	+ 188										
profit	1, S	+ 553								x		x
profit-sharing	1, 2e	- 361										
profiteer	1	+ 126										x
program	1	+ 616	x									

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
progress	1, 2s, E, S	+ 448	x	x
progressive		
progressivism	1	- .887	x
prohibit	1	+ .574	x
prohibition	1, 2g, S	+ .558	x
project	1	+ .258
proletariat	1, 2g	- .867
promenade	1	+ .085
prominence	1	+ .316
promiscuity	1	- .749
promise		
promissory note	1	- .074
promote	1, S	+ .632
promoter	1, 2e	+ .148
promulgate	1	- .163	x	x	x
pronounce	1	+ .558	x	x
proof	1	+ .558	x
propaganda	1, 2g, S	+ .206	x	x	x	x	x	x
propagate	1	+ .064
propertied class	1	- .059
property	1, 2s, S	+ .501
property tax	1	- .196	x
prophet	1	+ .332
proportionate	1	+ .127
proportional repre-		
sentation	1	- .928	x
proposal	1	+ .368	x
proposition	1	+ .368	x
proprietary article		
proprietary company	1	- .934
proprietary province	1	- .987	x
proprietor	1	+ .321
proscription		
prosecute	1, 2g, S	+ .322	x	x
prosecuting attorney	1, S	- .059	x	x
proselyting spirit	2s	
pro-slavery		
prosperity	1, S	+ .506
prostitute	1	+ .090
protagonist	1	- .928
protection	1, 2g, S	+ .506	x
protection of citizens		
abroad	2g	
protectionism	1	- .824

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
protective	1	— .250	x	.	x
protective tariff	1	— .771	x	x	.	x
protectorate	1, 2g	+ .500	x	x	x
protest	1	+ .052
Protestantism	1	— .861	x	x	x	.	x	.	.	x	.	x
protocol	1, 2g	+ .148	x
province	1	— .370	x	x	x	.	x
provincials	1	— .713
provincial government	1	+ .495	x
provisions	1, S	— .063
provisional government	1, S	+ .053	x	y
proviso	1	— .391
provocation	1	— .005	x	x
provoke	1	— 1.065
providence	1	+ .389
proximity	1	+ .327	x
psychology	1	— .005
psychopath	1	— 1.065
public	1, S	+ .389	y
public affairs	1	+ .327	x
public buildings	2e	— .005	x
public control of cable	2e	— 1.065
public control of power	2e	+ .389
public control of radio	2e	+ .327
public control of telegraph	2e	— .005
public control of telephone	2e	— 1.065
public defender	2e	+ .389
public domain	2e, 2g	+ .327
public education	2g	— .005
public health	1, 2g, 2e	+ .253
public health, quarantine	E	— .005
public industry	2e	— 1.065
public land	2g	+ .389
public law	2g	+ .327
public library	1	+ .558
public management	2e	— .005
public management and private ownership	2e	— 1.065
public management and public ownership	2e	+ .389
public officials	2e	— .005

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
public opinion	1, 2g, 2s, S	+ 252	x									x
public ownership	1, 2g, 2e	+ 132	x									
public property	1	+ 390	x									
public regulation of												
private enterprise	2e											
public school	1	+ 574										
public speaking	1	+ 453										
public spirit	1	+ 269	x									
public, the	2s, E											
public utility	1, 2g, 2e	- 233					x	x		x		x
public welfare	2g, 2e									x		
public works	2g, 2e									x		
publicity	1, 2g	+ 264										y
punish	S		x									
punishable	S		x									
punishment	1, 2g	+ 616	x				x			x		x
punitive expedition	1	- 644	x							x		
purchases												x
purchasing power	1	- 380										
pure democracy										x		
pure food law	1	+ 057	x				x	x				
puritan												x
puritanism	1	- 556										
quaesitor	1	-1 566	x									
quaker												x
qualifications	1, S	+ 385	x							x	x	
qualified veto	1	-1 019	x									
quantity theory	1	-1 082										
theory of												
money	2e									x		
quarantine	2g S								x			x
quarters (v)	1	- 392	x									x
quasi-public	1	- 772	x									
queen	1	- 632										x
queer			x									
quell	1	+ 148	x									
question (public)	S											
quibble	1	+ 026										
quit rent	1	-1 061										
quorum	2g, S		x							x		x
quota	1 2g	- 121	x	x	x		x					x
rabbi												x
rabble	1	- 027										x
race	1, 2s, S, E	+ 501	x									x

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Keity	Pressey
race balance	2s											
race conflict	1	-.054										
race inequality	2s											
race psyche	2s											
race suicide	1	-.387					x	x		x		
racial	1	+.099										
racial differences	2s											
racial minority	1	-.714	x									
racial prejudice	2s									x		
radiation	2s											
radicalism	1	-.527	x							x		z
radio												x
raid	1	+.379	x									x
railroad combination	2e						x					
railroaded									x			
railroad franchise	2e											
railroad rates, com-												
petitive	2e											
railroad rates, export	2e											
railways	S										x	
rajah	1	-.771	x								x	
rally	1, 2g	+.495	x									
rank	1	+.501	x								z	
ranks												x
ranking member	2g											
ransom	1	+.442										
rapacity	1	-.608										
rapine	1	-.766										
rapport	E											
rate (interest)	1, 2g, S	+.132									x	x
ratification	1, 2g, S	+.190	x							x		
ratify	S						x	x				z
ratio	1, S	+.195	x							x		
ration	1	+.258										
rationalism	1	-.935										
rationalization	2s, E									x		
ravage												x
raw materials	1	+.327								x		z
reaction	1, 2s, E	+.131	x						x			
reaction, circular	2s, E											
reactionaries	1	-.333	x	x						x		x
reactionism										x		
readings of a bill	2g									x		
readjustment	1, 2s	+.079							x			

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
ready-made	1	+ 495	.									
reaffirmation	1	- 201	.									
real estate	1, S	+ 495									jr.h.s	
realm	1	+ 199	x									x
real property	1	- 381	.									
reapportionment	1, 2g	- 375	x	x								
rebates	1	- 186	.									
rebel	1, 2g	+ 563	x									
rebellion	1, 2g	+ 506	x	x								
rebuff	1	+ 085	.									
recall	1, 2g, S	+ 085	x									
recalcitrant												
recant	1	- 433	.									
recapture												
receipt	1	+ 432	.						x		x	
receiver							x	x		x		
receivership	2c											
reception	1	+ 568	x								x	
recess	1	+ 553	x									
recession										x		
reciprocal	1	- 449	.									
reciprocity										x		
reciprocity treaty	1, 2e, 2g, 2s	- 777	x							x		
recite	1	+ 553	.									
reclamation	1, S	- 228	.							x		
recognition	1, 2g, 2s	+ 368	x						x		5	
recommendation	1	+ 626	x								x	x
recompense	1	+ 111	.									
reconciliation	1	+ 200	x							x		x
reconnaissance	1	- 166	.									
reconnoitre												
reconquest	1	+ 073	.									
reconstruction	1, 2g, 2s	+ 306	x									
records	1, 2g, S	+ 558	x								x	y
recreation	1, 2s, S	+ 568	.									x
recreational							x		
recreational facilities										x		
recruit	1	+ 443	x									
recuperation	1	+ .079	.									
recurrence												
redeem	1	+ 269	x									
redemption	1, 2g	+ 142	.									
redintegration	E		.							x		
rediscount	1, 2c	- 739	.									

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
redistribution	1	-.248
redoubt	1	x
redress	1	-.089	x
redress of grievances	1, 2g	-.507	x
"Reds"	1	-.523	x	x
red-tape	1	+.174	x	.	.	.	x	x
reduce	1	+.506	x	x	x
re-education	1	-.575
re-election	1, 2g, S	+.369	x	z	x
re-enact	1	+.206
reenforcements	1	+.311	z
referee	1	x	.	.
referendum	1, 2g, S	-.522	x	x	x	x	z
reforestation	1, 2g, 2e	-.052
reform	1	z
reformation	1, 2g, 2s, S	+.315	x
reformatory	1	+.385
reformer	1	+.326	x	.	.	.	x	x	.	.	z	.
refugee	1	+.011	x	x
refusal	1	+.632
regency	1	-.514	x	x	x
regicide	1	-.730	x	x	x	x	x	x	.	x	.	.
regimented	1	-.707	x
régime	1	-.566	x	z
regimen	1	-1.045
regiment	1	x
region	1	+.632	y
register	2g	x
Registration	1, S	+.316	x	x	x	.
regress	2s, E
regressive	1	x	.	.
regulars (army)	1	+.322	x	x
regularity	1	+.458	x
regulation	1, 2g, S	+.458	x	x	x
rehabilitate	1	-.259	x
reichstag	1	x
reign	1	+.385	x	x
reinstate	1	+.148	x	x
reinsurance	1	-.349
rejection	1	+.574	x	x
relapse	1	x	.	.
relations	1	+.637	x	x
relationships	2s, E, S	..	x
relatives	1	x	.	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No					Wesley College			
				I	II	III½	III	IV	Unpub	Pub	Kelty	Pressey
relax	1	+ 568									x	.
relics	1	+ 327										.
relief	1	+ 526	x							.	.	.
relief expedition	1	+ 327	x								.	.
religion	1, 2s	+ 437									.	x
religious dissemination	1	- 950									.	.
religious freedom	1	+ 073	x								.	.
religious laws	1									x	.	.
religious toleration	1	+ 011	x								.	.
relinquish	1	+ 127	x								.	.
remains	1	+ 685		
remand	1	- 798									.	.
remnant	1	+ 458									6	.
remodel	1	+ 579								.	x	.
remonstrance	1	+ 327	x	x			x				.	.
removal	2g										.	.
remunerative										x	.	.
renaissance	1	- 593								.	.	.
rendition	2g										.	.
renewal	1	+ 458	x								.	.
renomination	2g										.	.
renounce	1	+ 226	x								.	x
renovate	1	+ 269									.	.
renown	1	+ 342									.	.
rent	1, S	+ 379								x	.	.
renunciation	1	- 015									.	.
reorganization	E										.	.
reparations	1, 2g	- 300	x	x							.	x
repatriate	1	- 797	x								.	.
repeal	1, 2g, S	+ 516	x							x	.	x
repel	1	+ 453	x								.	.
repentance	1	+ 326									5	.
replacement	1	+ 232	x								.	.
replevin	2g									.	.	.
report	1, S	+ 569	x							.	.	.
report a bill	2g									.	.	.
representation	1, 2g, 2s, S	+ 390	x							.	x	.
representation, direct	.									x	.	.
representation, factional . . .										x	.	.
representation, minority										x	.	.
representation, political . .										x	.	.
representation, popular										.	x	.
representation, proportionate										x	.	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
									Unpub.	Pub.	Kelcy	Pressey
				I	II	III $\frac{1}{2}$	III	IV				
representative	2g, S	x	x
representative assembly	1	+.074	x
representative-at-large	x	.	.
representative government	1	+.153	x	.	.	.	x
repression	1, 2s, E	+.067	x
reprieve	2g	x	x	.	.
reprisals	1	-.397	x	x
reproduction	1	+.322
republic	1, 2g, S	+.310	x	x	x
Republican Party	2g	x
repudiate	1	+.127	x	x
request	1	+.626
requisition	1, 2g	+.016	x	.	.	.	x	x
rescind	x
research	1	-.116	x	.	.
resentment	1	+.442
reservations	1, 2g	+.057	x	x	x
reserve	1	+.322	x	x	.
reserved powers	x
reserve corps	1	-.379	x
reserve fund	1	-.323
reserves	2g, 2e, S	x	x	.	x
reserves, accumulated	x	.	.
re-shaping	2s
resident	1, 2g, S	+.558	x	x	x	.	x	x
residual	x	.	.
resignation	1, 2g	+.632	x
resistance	1, 2g	+.516	x	jr.h.s.	.	x
resolutions	1, 2g	+.553	x	x	x	x	.	x	.	.	x	x
resources	1, S	+.443	x	x	.	x
respect	1	+.569	x
response	2s, E
responsible government	2g
responsible ministry	1	-.887	x	x	.	.
responsibility	1, 2s, S	+.385	x	x	x	x	x	x	.	.	x	.
restaurant	1	+.747
restitution	1	-.445
restlessness	2s, E
restoration	1	+.183	x	x	.	x
restraint of trade	1	-.634
restriction	1	+.385	x
result	4	.
resumption	1	+.127	x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
resurrection	1	+ .211
retail	1	+ .258
retail prices	1	+ .267
retailer	1	+ .191	x
retainer	1	+ .072
retaliation	1	+ .378
retardation	2s	+ .572
retirement	1	+ .398	x
retreat	1	+ .369	x	x
retrenchment	1	- .362
retribution	1, 2s	- .791	x	x
retrieve	1	- .640
retroactive	1	- .582
retrograde	1	- .511	x
retrogress	1	+ .061
returns	2g	+ .073
reunion	1	+ .578
revaluation	1	+ .328	x	x
revelation	1, 2s	+ .328	x	x
revenge	1, 2s	+ .328	x	x
revenue	1, 2g, S	+ .328	x	x
revenue, anticipatory	2e	- .263	x	x
revenue, internal	1, 2e, 2g	+ .323
reverence	1, 2s	+ .184	x
reverses	1	+ .184	x
revision	1, S	+ .183
revival	1, 2s	- .325	x
revocation	1	+ .395	x	x
revolt	1, 2g, 2s, S	+ .395	x	x
revolution	1, 2g, 2s, S	+ .395	x	x
revolutionary	1, 2g, 2s, S	+ .395	x	x
revulsion	1	+ .573	x
reward	1	- .781	x	x
rider (leg.)	1	+ .439
ridicule	1	- .898	x
"Right" (n. pol.)	1, S	- .981	x
right of belligerents	1	- 1.003
right of deposit	1	- .897	x
right of neutrals	1	- .742	x
right of recall	1	- .742	x
right of search	1	- .742	x
right of way	2e	- .742	x
right classes	2e	- .742	x
ring (pol.)	S	- .742	x

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
ringleader	1	+.389	x
riot	1	+ .517	x	z
riparian rights			x	.	.
risk	1, 2e	+.573	x	x	.	x	.	.
rite	x
ritual	1 2s	-.078
rivalry	1, 2g, 2s, E	+.451	x	x
road surveyor	2g	
road tax	1	-1.003	x
robbery	1	+.573
roll-call	1, 2g	+.5556	x	x
rolling stock	x	.	.
romance	x	.	.
romantic	x	.	.
rostrum	1	-.286	x
rotation	1	+.123	x	x	.
rout	1	+.139	x
routes	1	+.506	x	x
rowdies	1	+.389	jr.h.s.	.	.
royal	1	+.456	x	x
royal ceremonies	x	.	.
royalists	1	+.043	x
royalty	x	.	x
rudiments	1	-.096
ruins	1	+.573
rule of merit	x	.	.
"rule of reason"	2g	x	.	.
ruler	1	+.573	x	x	.	x
rules	S
rum running	1	+.456
rural	1, S	+.389
rural free delivery	2g	z
ruthless	1	+.206	x
sabotage	1 2e	-.815	x	.	x
sack	1	+.328	x
sacraments	1	-.637
sacred	1	+.328
sacrifice	1	+.389
sacrilege	1	-.341
safe deposit box	1	+.511
safeguard	1	+.328	x	x
sahib	x	.	x
safety legislation	2e
sailor	x

APPENDIX II

Word	List Symbol	Morton Weights	Wealey List	Wealey Tests, Editions No.					Wealey College			
									Unpub	Pub	Ketty	Pressay
				I	II	III½	III	IV				
saint	1	+ 328										
salaries	1, S	+ 611	x									x
sale	1	+ 573							x			x
salesman	1, 2e	+ 573					x					
salestax	1	- 478	x									
salient	1	- 663	x	x								
saloon	1	+ 328										x
salute												x
salvation	1	+ 328										x
salvo												
Samurai	1	- 1 314										
sanction	1, 2s	- 169										22
sanctity	2s											
sanctuary	1	- 563										
sane	1	+ 517										
sanitation	1, S	+ 317									2	
satrap	1	- 1 071	x	x								
satellite	1	- 753	x	x								
saturated										x		
savages	1	+ 578										x
savant	1	- 920										
savings	1, S	+ 573								x	2	
scabbard												x
scab	1, 2e	- 157					x	x		x		y
scalawag												x
scandal	1	+ 511										
scarcity	1	+ 451										
scepter	1	- 133	x									
schedule	1	+ 450	x								x	
scheme	1	+ 511										
schism	1	- 793										
scholar	1	+ 311										
scholarship	1	+ 368										
scholasticism	1	- 1 108										
schools	1, S	+ 672										x
school board	1	+ 558									x	
science	1	+ 253									x	
scientific management	1, 2e	- 597										
scientific research	2s											
scolding	1	+ 689										
scout												x
scutage	1	- 1 429	x	x								
sea-level												x
sealed legal											x	

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
sealer of weights and measures	2g
seaport	1	+ .384	x
sea-power	1	+ .095	x
seasonal	1	+ .043
search-warrant	1	- .083	x	x	.
seat	x
secede	1, 2g	+ .264	x	y	z
seclusion	1	+ .326
secondary schools	1	+ .379
second-rate power	1	- .287	x
secrecy	2s
secret ballot	1	- .214	x
secret diplomacy	1	- .529	x
secret service	1, 2g	- .052	x	x
secret society	1	+ .379	x	.	.
secretariat	1	- .718	x
secretary	2g, S	.	x	z
sect	1	- .143	6	x
sectarianism	1 2s	- .627	x
section	S	.	x	y
sectionalism	1, 2s, S	- .297	x	x
sector	1	- .289	x
secular	1	- .582	x
securities	1	- .058
security	1, 2s	+ .390	x	x
sedition	1, 2g, 2s, S	- .333	x	x	z
see (papal)	x
segregation	1, 2s, E	- .432	x	.	.
seignorage	2e	x	.	.
seize	1	+ .689	x
selection	2s, E, S
selective draft	1	- .381	x
selective exclusion	x	.	.
selectman	1, 2g	- .676	x	x	x	.	x	x
self	E
self-control	1	+ .390	5	.
self-conscious behavior	2s
self-defense	1	+ .448
self-denial	1	+ .385	x	.
self-determination	1, 2g	- .559	x	x
self-education	1	- .017
self-government	1, 2s, 2g	+ .006	x	x	z
self, expanded	2s

APPENDIX II

Word	List Symbol	Morton Weights	Weeley List	Weeley Tests, Editions No.					Weeley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Preuey
self, mirrored	2s											
self preservation	1	+ .195										
self reliance	1	+ .201									6	
self sacrifice	2s											
self sufficient	1	- .226										
self sufficient nation	2c											
self support	1	+ .327										
sell	1	+ .747										
senate	1, 2g, S	+ .322	x								5	2
senator	2g, S											2
senatorial courtesy									x		
seniorage									x		
seniority rule	2g									x		
sensation	1	+ .437									jr.h.s.	
sensuality	1, 2s	- .122										
sentence	1	+ .558	x								x	x
sentiment	1	+ .253	x									
sentinel	1	+ .443	x				x	x				x
separation	1	+ .558								x		
separatists											x
sepoys	1	-1 114	x	x	x	x	x	x		x		x
sequester											x
seraglio	1	-1.476										
serf	1	- .528										x
serfdom	1	- .528										
sergeant	1	+ .237	x									x
sergeant-at-arms	2g, S		x							x		
sermon	1	+ .627									2	
servitude											x
servant	1	+ .616										x
service (army)											x
services	1, S	+ .458	x									x
servile											x
sesame	1	-1.050										2
session									x		
sessions	1, 2g, S	- .269	x	x								
session, long	1, 2g		x									
session, short	1, 2g		x									
setback	1	+ .342	x									
settlement											2
settlement house	1	- .312										
settlements	1, S	+ .427	x									
sever	1	+ .274									jr.h.s.	
sex	1	+ .574										

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
shackles	1	+.211
shame	1	+.616
shareholder	1, 2e	-.047
shares	x	.	x
sheriff	1, 2g, S	+.358	x	.	.	.	x	x	.	.	x	x
shield	1	+.458	x
shilling	x
ship-building	1	+.569
shipment	1	+.532
ship-money	1	-1.065	x
ship owner	1	+.385
ship-subsidy	1	-.893
shipping	1	+.269	y
shipping company	x	.	.
shire	1	-.625	x
shogun	1	-1.382	x	x
shop	2e
shop closed	2e
shop, committee	2e
shopkeepers	1	+.579
shopping goods	x	.	.
shop, open	2e
short ballot	1	-.798	x	x	.	.
show	1	+.600
shrine	1	+.137
siege	1	+.458	x	x	x
signal	1	+.495	x
signature	1	+.632	x	x	x
significant	E
signor	x
silver, bullion	2e
silver certificate	2e
silver dollars	2e
simian	x	.	.
simony	1	x	.	.
simulation	2s, E
sinecures	1
single tax	1, 2e, S	x	x
sinking fund	1, 2e	x	.	.
sire	1
site	1	x
situation	E
skepticism	1
skill	1

APPENDIX II

Word	List Symbol	Morton Weights	Wealey List	Wealey Tests, Editions No.					Wealey College			
				I	II	III½	III	IV	Unpub.	Pub	Kelcy	Prescy
skipper	1											
skirmish	1		x									x
slackers	1		x									
slander	1		x							x		
slaughter	1		x									
slave												
slave trade	1		x									x
slave holder												x
slavery	2s		x									x
slay	1											
slogan	1		x									x
slums	1											x
small farmer	1											
smuggling	1		x	x	x		x	x		x		x
snobbery	1											
sociability	2s								x			
social	E, S											x
social achievement	2s											
social activity	2s											
social adaptation	1											
social adjustment	2s											
social aggregation	2s											
social ascendancy	2s											
social assimilation	2s											
social attitude	2s											
social authority	2s											
social automatism	2s								x			
social balance	2s											
social case work									x			
social center	S											
social change	2s											
social choice	2s											
social control	1, 2s									x		
social discontent	1											
social disintegration	2s											
social distribution	2e											
social equality	1											
social evils									x			
social evolution	1, 2s											
social feeling	2s											
social fermentation	2s											
social flexibility	2s											
social genesis	2s											
social group, primary	2s											

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
social groups, secondary	2s
social heritage	2s	x	.	.	.
social instability	2s
social instinct	2s
social institution	1, 2s	x	.	.
social insurance	1
social integration	2s
social interest	2s
social justice	1, 2s
social lags	x	.	.
social life	2s
social manipulation	2s
social mechanism	2s
social mind	2s
social mobility	1
social movement	2s
social nexus	x	.	.
social order	2s
social organism	2s	x	.	.
social organization	2s
social ossification	2s
social pathology	1, 2s	x	.	.
social pattern	1, 2s	x	.	.
social polity	2s
social position	1
social pressure	2s
social problem	2s
social processes	1, 2s
social product	2s
social progress	2s, 2c
social psychology	1	x	.	.
social purposes	2s
social reactions	2s
social reform	1, 2s
social regeneration	2s
social regulation	2s
social relationship	2s
social ritual	2s
social sanction	2s
social science	1
social selection	1, 2s
social self-consciousness	2s
social service	x
social situation	2s

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelcy	Pressey
social spirit	2s
social stability	1
social stagnation	2s
social status	2s
social stratification	2s
social structure	2s
social suggestion	2s
social surplus	2e
social sympathy	2s
social trend	2s
social type	2s
social uniformities	x	.	.
social unity	2s
social unrest	2s
social valuation	2s
social wants	2s
social waste	2s
social well-being	2s
social will	2s
social worker	1
socialism	1, 2g, 2e	x
socialist	S	z
socialist party	2g
socialization	2s, E
societal	E
society	1, 2s, E, S	z
society, civil	2s
society column	1
society, feudal	2s
society, pastoral	2, 2s
society, political	2s
society, primitive	2s
society, static	x	x	.	x	.	.
society, sympathetic	2s, E
society, tribal	2s
sociocracy	2s, E
sociological research	x	.	.	.
sociology	1, 2s, E
socio-mechanisms	x	.	.
sociosphere	E
socius	2s, E
soldiers	1	...	x	x	z
solicitation	1
solicitor	1, 2g, S	..	x	x

TESTS AND MEASUREMENTS

Word	Last Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
									Unpub.	Pub.	Kelty	Pressey
				I	II	II½	III	IV				
"Solid South"	2g		x
solidarity	1, 2s, E	..	x	x
solidified									.	x	.	.
solicitude	2s	
solvent		x
sophistry	1	
sorority	1	
sound money	1	
source	1, S
source documents		
sovereign	1, S	...	x	x	x	.	.
sovereign people	2g	
sovereign power	2g	
sovereign state	2g		x	.	.
sovereignty	1, 2g, 2s, S		x	x	.	.
soviets	1, S		x	x
spare time	1	
speaker	2g, S		x	.	x
speakeasy			x
specialist	1	
specialization	1, 2s	
special interests	1	
special legislation	1	..	x
special privilege	1		x
specialty goods			x	.	.
specialty shop	1	
specie	1		x
specie payment	x	.	.
specific			x	.	.
specific duty	1		x
spectacle	1	
speculation	1, 2c, S	y
speculator	2c
spendthrift	1	
sphere of influence	1, 2g	.	x	x	.	.
spies	1		x
spirit	S
spiritism		x	.	.
spiritual	1
spiritual qualities	x	.	.
splendor	1
split	x
spoilsman	1	x
spoils system	1, 2g, S	x	z

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub	Pub.	Kelty	Pressy
spokesman	1		x									x
sport										x	.	.
spurn	1										.	.
spurious value									x		.	.
spy	1		x								.	.
squabble	1								.		.	.
squad											.	x
squadron	1		x								.	x
squalor	1										.	.
square deal	1		x								.	.
squatter											.	x
squire	1										.	x
stable										x	.	.
stable currency	1										.	.
stability	1		x						x		.	.
stabilization	1	- 386									.	x
stabilized dollar	2e										.	.
stadt-holder	1	-1 266	x								.	.
staff	1	+ 269	x								.	x
stagecoach											.	x
stage of culture	1	- 591									.	.
stagnation	1, E	+ 269									.	.
stake	1	+ 495									.	.
stamp	1	+ 558									.	.
stamp tax	2g										.	.
standard	1 2e, S	+ 269									.	.
standard, gold	2e										.	.
standard of conduct	1	- 132									.	.
standard of living	1, 2e, 2s	- 314									.	z
standardized dollar	2e										.	.
standardized tests									x		.	.
standing army	1	+ 322	x								.	.
stand-patters										x	.	x
staple	1	- 074									y	x
starvation	1	+ 569									.	.
state	1, 2g, S	+ 448	x							x	.	z
state aid	1	- 307	x								.	.
state bank	1	- 281									.	.
statecraft	1	- 385	x							.	.	.
state equality	2g							
statehood	S									.	.	.
state (of flux)	1	- 548								.	.	.
state of war	2g	+ 0005	x							.	.	.
state paper	1	- 375	x							.	.	.

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
state's evidence	2g	x	.	.
statesman	1, 2g, S	+ .195	x	y
statement (bank)	1	- .005	x	.
state religion	1	- .466	x	x
state rights	1	- .481	x	z
state socialism	1	-1.019	x
the state	2s
station	1	+ .616	z	.
stationary	1	+ .505
stationary population	x	.	.
statistical research	x	.	.
statistics	1	+ .079
status	1	+ .022	x
status quo	1, 2g	- .592	x	x	x	x	x	x	.	x	.	.
status quo ante bellum	x	x	.	.	.
statute	1, 2g, S	+ .188	x	x	x	z	.
stay	x	.
steadiness of endeavor	2s
steal	1	+ .616
steerage	1	- .122
steering committee	2g	x	.	x
sterilization	1	- .382
sterling	x
steward	1	+ .127
stifle	1	+ .258
stimulation	1, 2s, E	+ .327
stockade	1	- .006	x	x
stock company	1, 2e	- .593
stock exchange	1, 2e	- .323	x	x
stock holder	x
stocks	S	x	.	z
stoppage at source	2e
storage	2e
store	1	+ .690	z	.
stories	1	+ .616
straits	x
stranger	2s, E
strategic	z
strategy	1	+ .126	x	x
stratification, social	2s, E	x	x	.	x	.	.
straw vote	x
strict construction	1	- .572	x	x	x	.	.
strife	1	+ .506	x
strike	1, 2e, S	+ .506	x	z

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III 1/4	III	IV	Unpub.	Pub.	Kelty	Pressey
strike-breaker	1, 2e	+ .062
stringent	1	- .301	x
stronghold	1	+ .327	x
structural reforms	1	x	.	.
structures	1	x	.	.
struggle	1, 2a, E	+ .495	x	x
stubbornness	1	x	.	.	.
students	1	+ .689
stump speaker	1	+ .022	x
sub-deacon	1	- .090
subdivision	1	+ .206	x
subdue	1	+ .443	x	y	.
subjects	x
subject states	1	- .522	x	x
subjugation	1, 2a, E	+ .083	x	x
sublimation	E
submarine	22
submission	1	+ .269	x
subnormal	1	+ .020
subordinates	1	+ .211	x
subpoena	2g	x	x	.	.	x	x	.	x	.	x
subscription	1	+ .437
subservience	1, 2a	- .231	x
subsidiary coin	2e
subsidiary	1	- .347	x
subsidies	1, 2g, 2e	- .363	x	x	x	.	x	x	.	x	.	.
substitution	1	+ .363
sub-treasury	1	- .523	x
suburb	1	+ .311	x
sub-vaasal	1	- .754	x
success	1, 2a, S	+ .553	x	4	.
succession	1	+ .385	x
succor	1	- .175	x
sue	1	+ .442
suffrage	1, 2g, S	+ .211	x	x	2
suffragette	1	+ .132	x	x
suggestion, social	2a, E
suggestibility	E
suit	1, S	+ .379	x	x
suitor	1	+ .327
sulphurous	x	.	.
sultan	1	- .047	x
summation	E
summons	2g	x

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelsey	Pressey
summonees										x		
superficiality	1	-.301								x		
sumptuary law										x		
Sunday laws									x			
superintendent	1, 2g, S	+.437	x									x
superiority	1	+.495	x								jr.h.s.	
supernatural	1	-.005										
super-organic	E											
superstition	1	+.311								x		
super-tax	1	-.977	x									
supervision	1	+.316	x									
supervisor	2g, S		x									
supplant	1	+.216										
supplies	1	+.558	x									x
supply and demand	1	-.609										
support	S		x									x
supporter	1	+.632	x	x						x		
suppression	1, 2s, E	+.458	x									x
supremacy	1	+.458	x									x
Supreme Court	1, 2g, S	+.316	x									z
surety	1	-.486										
surpass	1	+.400										
surplice and ring	1	-1.139										
surplus	1, 2e	+.326							x		x	x
surplus, accumulated							x			
surplus earnings							x			
surplus value	1	-.987										
surplus wealth							x			
surrender	1	+.690	x									z
surrogate	2g										
surroundings	S										
sur-tax	1, 2e	-1.192	x									x
surveyor	2g										
survival	1, 2s, E	+.316										
survival of the fittest	1, 2s, E	+.316								x		
suspects										x
suspend	1	+.569	x	x								
suspension								x		x
suspensive veto	1	-.929	x									
suspicion	1	+.574									5	
sustaining group	2s										
suzerain	1	-.993	x									x
swashbuckling								x		
sway	1	+.400	x									

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III 1/2	III	IV	Unpub	Pub.	Kelcy	Pressey
sweat shop	1, 2e, S	- 382					x					x
swindle	1	+ 448										x
symbol	1, 2a, E	+ 137	x									
sympathetic strike	1	- 493										
syndicalism	1, 2e	- 1 024	x									
syndicate	1	- .571										x
synergy	E											
synod	1	- 987										x
system	1, S	+ 437	x									
synthesis, social	E											
table (v.)	1, S	+ 501	x									x
taboo	1	- 465								x		
tactics	1	- 269	x									x
taille	1	- 1 398	x									
tales												
talesman	1, S	- 958							x			x
talisman	1	- 861	x						x			
tame	1	+ 632										
tangible investment										x		
tariff	1, 2g, S	+ 327	x				x			x		x
tariff commission	2g											
tariff duties	1	+ 194										
tariff for revenue	1	- 259	x									
tariff preferential	2e											
tariff, protective	2e, 2g											
tariff rates	2g											
tariff schedule	2g											
tariff treaty										x		
tariff union	1, 2e, 2g	- 703										
tax	2g, S											x
taxable taxpayer	S											
taxation	1, 2g, S	+ 558	x									
tax-dodger	1	+ 241	x									
taxfree minimum	2e											
taxpayer	1, 2g	+ 747	x									
tax rates										x		
Taylor System	2e											
teacher	1	+ 747										
teamwork	1, S	+ .747									x	
technicality	1	+ 074										
technical school	1	+ 253										
technique, social	E							x				
telesis	2s, E											x
telegraph												

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelcy	Pressey
telephone												zz
temperance movement	1	+.385					x	x				
temperament	2s											
temples	1	+.390										
temporal	1	-.371										
temporal power	1	-.656	x									
temptation	1	+.558									6	
tenant	1	+.448								x		
tenant farmer												x
tenant-in-chief	1	-.986										
tendency	1, 2s	+.390										x
tender												x
tenement	1, S	+.206										
tension	E											
tenure	1	-.444	x									
tenure of employment										x		
tenure of office	2g		x									
term (office)	1, 2g, S	+.616									x	x
termination	1	+.269										
terminal	1	+.269										
terms (treaty)	1	+.253	x								x	x
territorial expansion	1, S	+.069	x									
territorial integrity	1	-.862	x	x	x	x	x	x				
territory	1 2g, S	+.443	x							x		zz
terror	1	+.622	x	x							4	
terrorism	1	-.236	x									
test	1	+.558	x									
testator												x
test oath	1	-.745	x									
testify												x
testimony	2g		x									z
textile												x
theft	1	+.616										
theology	1	-.408										x
theocracy	1, 2g	-.934	x	x	x	x	x					
theory	1	+.132	x									
therapeutics										x		
they-group	2s, E											
third degree	1	-.396										
third estate	1	-.951	x									
third party	1, 2g	-.260	x									
threaten	1	+.689	x								5	
thrift	1, S	+.521									z	
throne	1	+.579	x									x

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
thwart	1	+ 279										
ticker												x
ticket	2g, S											x
tillage												x
time and motion studies										x		
tithe	1	- 582										
title	1, S	+ 132									x	x
title deeds	1	- 583										
tocain	1	-1 003										
toga	1	- 703	x									
toleration	1, 2s	-1 387										x
toll	1, S	+ 063	x									x
tomahawk												x
tomb	1	+ 579										
tonnage	1, 2g	+ 006										x
tools										x		
torch bearer	1	- 110										
torics	1	- 407									jr.h.s.	
torment	1	+ 579										
torpedo	1	- 374	x				x					x
tort	1, 2g	-1 066	x	x								
torture	1	+ 264										
totem									x			
totemism	2s											
tournament	1	+ 127										
town	2g, S											22
town meeting	1, 2g, S	- 068	x									
township	1, 2g, S	- 025	x								x	x
tract	1	+ 243										x
trade	1, S	+ 632										2
trade agreement	1	- 324										
trade area										x		
trade association	1	- 382										
trade guilds	1	- 629										
trade, interstate	2e						x	x				
trade, intrastate	2e											
trademark	2g											
trade monopolies	1	- 508										
trade regulation	1	- 113										
trade route	1	+ 126									x	x
trade school	1	+ 316										x
tradesman												
trade union	1, S	+ 258										
trading post	1	+ 074										

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests Editions No					Wesley College			
									Unpub	Pub	Kelty	Pressey
				I	II	III 1/2	III	IV				
tradition	1, 2s, f	+ 269	x									x
traffic	1	+ 564	.									x
traffic association	2e	.	.									
training	S									x		
trait	1	+ 269									x	
traitor	1	+ 448	x									x
tranquillity	1	- 042	x									
transaction												x
transcontinental	1	+ 379										x
transfer	1	+ 448										
transference of ideas									x			
transformation	1, 2s, E											
transit	1	+ 079										x
translation	1	+ 443									x	
transmission of acquired characteristics												
transmutation	2s									x		
transport												x
transport agency									x			z
transportation	1, S	+ 443							x			
transvaluation										x		
transubstantiation	1	- 851										
travel	1	+ 616										x
traveler's check	1	- 476										
treachery	1	+ 369										
treason	1, 2g, S	+ 258	x				x	x		x		z
treasure	S											
treasurer	2g, S		.									
treasury	1, 2g, S	+ 369	x								x	zz
treasury notes	2g, 2e											
treat (v.)	1	+ 336	x									
treaty	1, 2g, S	+ 258	x							x	z	z
treaty of alliance										x		
treaty ports	1, 2g	- 708								x		
treaty zone			.						.	x		
trench											.	x
trend	1	+ .090	.			.					.	
trespass	2g								.	.	.	
trial	1, 2g	+ .563	x	x
trial by combat	1	- 765	x			
trial by jury	1 ..								.	x	x	.
tribe	1	+ .501	x	.								x
tribunal	1	- .333	x	.						.		x
tribune	1	- .756	x	x			.		.	.		

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelly	Pressey
tributary	1	- 137	x									
tribute	1	+ 269	x									x
trickery	1	+ 443										x
tricolor	1	- 818	x									
tripartite												
triumph	1	+ 506										x
triumvir											5	x
triumvirate	1	- 729	x	x								x
troops	1	+ 574	x									x
tropics												x
trouble	1	+ 747	x									x
trover	2g											
truancy	1	+ 500										
truce	1	+ 327	x				x	x				x
trunk lines	1, 2e	- 317										
trust	1, 2e, S	+ 046										
trust busting	1	- 639								x	je h e z z	
trust company	1, 2e	- 724										
trustees	1, 2e, 2g,	- 012					x	x				x
trust fund	2e											
tsar	1	- 021	x									
tuition	1	+ 558										
turbulence	1	+ 153										
turmoil	1	+ 269	x									
tutor	1	+ 326										
twelve-hour day	1	+ 132										
two-party system	1	- 408	x	x								
type	1	+ 379										
typical												
tyranny	1, 2g, S	+ 199	x							x		2
tyrant												x
ukase	1	- 140	x	x						x		x
ultimatum	1, 2g	- 192	x	x						x		x
ultra-conservative	1	- 592	x									
umpire	1	+ 432										
unadjustment	E											
unanimity	2g											
unanimous	1	+ 579	x				x	x				2
unconstitutional	1, 2g, S	- 011	x							x		2
underconsumption	2e											
undercut												x
understand	1	+ 558										
underworld												x
underwriting	2e								x			

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
underwriter	x
unearned increment	1, 2e	-.982
unemployment	1, 2e, S	+.448	x	.	.
unemployment, cyclical	2e	x	.	.	x	.	x
unemployment insur-
ance	1, 2e	-.519	x	.	.	.
unemployment, seasonal	2e
unfair lists	1	-.719
unfit	1	+.517
unification	1	+.295	x
uniformity	1	+.269	x
union	1, 2g, S	+.443	x	4	x
unionize	1	-.047
unique	1	+.153	x	.	x
unit	1	+.269	x
unit rule	2g	x	.
unite	1	+.437
United Press	S
United States	S
United States treasury	x	.	.
unity	x	.	.	.
universal	1	+.264	x
universal education	1	+.206
universal military
service	1	+.019	x
universal pattern	E
universal suffrage	1	+.089	x
universe of discourse	E
university	1	+.501	x
unjust	1	+.564	x	x	x	.	x
unlawful	x	x
unofficial	1	+.506	x
unorganized	1	+.390	x
unorthodox	x	.	.
unorganized labor	1	+.132
unprofitable	x
unrepresentative	S
unrest	1, 2s	+.322	x
unscrupulousness	1	+.079
unskilled labor	1	+.390	x	x
unsocial	1	-.138
unsound money	x	.	.	.
untaxed wealth	x	.	.
untrustworthy	x	.	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
unveiling	1	+ 190										
upheaval	1	+ 327	x									
upholder	1	+ 190	x									
upper class	1	+ 257										
upper house	1	+ 178	x									
uprising	1	+ 448	x									x
upstart	1	+ .211										
urban												
urban centers	1, S	+ .003										
urbanization	2s									x		
urban population	1	+ 003										
usages	1	+ 148										
usurpation	1	- 297	x									
usury										x		
utilitarian	1	- 814										
utilities	1, S	- 454										x
utility	2e						x	x		x		
utilization	1, 2s, E	- 047										
utopians	2e		x									
vacancy	1	+ 264									6	
vacation	1, S	+ 747										x
vagrant	1, S	+ .148	x									x
valid	1, S	+ 143	x							x		x
valor	1	+ 485	x								6	
valuation	1, 2e, 2s, E, S	+ 264	x							x		x
value	2e											
vandalism	1	- 445	x	x								
vanguard	1	- 134	x									
vanquish	1	+ 327									z	x
variability										x		
variation	2s, E									x		
variety	1	+ 500										
vassal	1	- 454	x									
vassalage										x		
vault	1	+ 327										
velocity of circulation	2e											
venal												x
vener	1	- 107										
veneration	1	+ 148										
vengeance	1, 2s	+ .442									jr.h.s.	
venire	S									x		x
venture	1	+ .269							x			
veracity	1	- 054										
verdict	1, 2g	+ .269	x	x	x			x		x		z

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Pressey
verification	x	.	.
vernacular	1	-.507	x	.	.
Versailles Treaty	S	
verticle combination	1	-1.313	x	.	.
vessel	x
vested interests	1, S	-.950	x	.	x
veteran	1	+.443	x	z
veto	1, 2g	+.495	x	x	z
veto item	x	.	.
vexation	1	+.343
vice	1	+.385	x	.	.
vice and crime ring	x	.	.
vice-president	1, 2g	+.437	x	x	zz
viceroy	1	-.634	x
vice-royalty	1	+.495	x
victims	1	+.495
victory	1	+.632	x	y
villa	1	-.260
village	1, 2g, S	+.422	z	x
villain	1	-.259
villains	1	-.270
vindictiveness	1	-.281
violation	1	+.258	x	x	.	.	z
violence	1	+.553	x
virtue	1	-.385	x	.
visa	1, 2g	-.708	x
visé	S	x	.	.
viscount	1	-.861	x
visitation	1	-.044
visiting nurse	1	+.211	x
vital statistics	1, 2g	-.508	x	.	.
viva voce	2g
vizier	1	-1.053
vocation	x
vocational guidance	1, S	-.387
void	1, 2g	-.216	x	x
volition	2s
volley	x
Volstead Act	S
voluntary arbitration	1	-.369
volunteer	1, 2g	+.443	x	z	z
volunteer army	1	+.322	x
volunteer clubs	x	.	.
vote	1, 2g, S	+.506	x	x	x	z

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III $\frac{1}{2}$	III	IV	Unpub.	Pub.	Kelty	Pressey
voting machine	2g
voting unit
vow	1	+.327	6	.
voyage	zz
vulnerability	1	-.364	x
wage	1	+.564	x	.	x
wage earner	1	+.448	5	.
wage scale	1	-.100
walk-out	1	-.296
wants	1, 2s, E, S	+.632
war	1, 2g, S	+.616	x	z	y
war cloud	1	+.132	x
war debt	1, 2g	+.506	x
ward off	1	-.312	x
ward	1, 2g, S	+.269	x	z	x
warden	1, 2g	+.132	x	x
warehouse	1	+.327	x	.
warehouse receipt	2e
warfare	x
warning	1	+.569	x
war power	2g
warship	x
warrant	1, 2g, S	+.137	x	x	.	x
warrior	x
Washington Conference	S
waste	1, S	+.437	x	.
wasteful consumption	2e
wastes of competition	2e
watcher	2g, S	x
watchful waiting	1	-.713	x	x
watered stock	1	-.828	x
waterpower	1, 2e	-.099	z	.
waterpower sites	2g
waterway	1, S	+.143	x
Ways and Means
Committee	2g
wealth	1, S	+.374	x	4	z
weapons	1	+.632	x	x	.	x
weather bureau	1, 2g	+.390
we-feeling	2s
we-group	E
welfare	1, S	+.327	x	x
welfare worker	1	+.074
well-to-do	1	+.390

TESTS AND MEASUREMENTS

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	II½	III	IV	Unpub.	Pub.	Kelty	Pressey
wergeld	1	-1.018	x
westernization	1	-.777	x
Westward Movement	1	-.339	x
wet	x
"what the traffic will bear"	2e
Whigs	1	-.966	x	x
whip	S	x	.	x
whipping post	x
whiskey	x	.	.	.
white man's burden	1	-.987	x	x	x	x
White House	x
wholesale	1	+.374	jr.h.s.	.	x
wholesome	1	+.385	6	.	.
wield	1	+.201	x
wilderness	4	.
will	1, S	+.501	x	x
will, social	E
wish	E
witan	1	-1.424	x
witchcraft	1	-.175	x
withdraw	1	+.501	x	x
witness	1, 2g, S	+.448	x	x	z
woman's suffrage	1, S	+.263	x
work	S
working conditions	1, 2e	+.085	x	.	.
workman's compensa-
tion law	1, S	-.751	x	.	.
world	S
world court	1	-.306	x	x
world despotism	1	-.851	x
world empire	1	-.403	x
world monarchy	1	-.582	x	x
world peace	1	-.032	x	6	.
world power	1, 2g	-.259	x	x	x	.	.
worship	1	+.385	z	.
worth	1	+.443
wounded	x
wrangle	1	+.206
wreck	4	.
wrest	1	+.211	x
writ	1, 2g, S	-.302	x	x
writings	1	+.385
written constitution	1	.000	x	z	.

APPENDIX II

Word	List Symbol	Morton Weights	Wesley List	Wesley Tests, Editions No.					Wesley College			
				I	II	III½	III	IV	Unpub.	Pub.	Kelty	Presary.
written law	1	- 148	x	x
wrong doing	1	+ 327
yellow dog contract
"yellow" newspapers	1	- 402
yeomanry	1	- 919
yield	1	+ 579
yoke	1	+ .326	x	y	.
Young Plan	2g
zeal	1	- .384
zealot	x
Zeppelin	x
zone	1, S	+ 327	x	.	.	.	x	.	.	x jr.	h.s.	x
zoomorphism	2s

BIBLIOGRAPHY OF TESTS

AND

CRITICAL DISCUSSIONS OF TESTS

History

- American Council Ancient History Test*, Raymond, I. W., Timmerman, H., Wood, B. D. High schools and colleges. World Book Co., New York, 1928.
- American Council European History Test*, Carman, H. J., Langsam, W. D., Wood, B. D. High schools and colleges. World Book Co., New York, 1928.
- American History*. Harlow Publishing Co., Oklahoma City, Okla. High school.
- Analytical Scales of Attainment in American History*. Educational Test Bureau, Minneapolis, Minnesota. Grades 7 to 8.
- Ancient and Medieval History*. Harlow Publishing Co., Oklahoma City, Okla. High school.
- Barnard, A. F., *Test in Roman History*. University High School, University of Chicago.
- Barr, A. S., *Diagnostic Tests in American History*. Public School Publishing Co., Bloomington, Ill. 1920.
- Barr-Daggett Tests in American History*. Educational Test Bureau, Minneapolis, Minn. Senior high school.
- Barr, A. S., Gifford, C. W., *American History Word Meaning Test*. Public School Publishing Co., Bloomington, Ill. 1931.
- Bell, J. Carleton, and McCollum, D. F., "A study of the attainments of pupils in United States history," *Journal of Educational Psychology*, VIII (May, 1917) 257-274.
- Bowman United States History Test*. Public School Publishing Co., Bloomington, Ill.
- Brooks, Fowler D., and Basset, S. Janet, "The retention of American history in the junior high school," *Journal of Educational Research*, XVIII (October, 1928) 194-202.
- Broome, E. C., and Boyer, P. A., "Report on survey test in history of the United States, grade 7B," *Historical Outlook*, XX (January, 1929) 17-23.
- Buckingham, B. R., "A proposed index of efficiency in teaching United States history," *Journal of Educational Research*, I (March, 1920) 161-171.

BIBLIOGRAPHY OF TESTS

- Buckingham, B. R., "The correlation between the ability to think and the ability to remember, with special reference to United States history," *School and Society*, V (April 14, 1917) 443-449.
- Camp, F. S., "Wanted: a history scale maker," *Journal of Educational Research*, II (June, 1920) 517-518.
- Cavins, L. V., "An experiment with standardized tests in a state teacher's examination," *Journal of Educational Research*, XIV (October, 1926) 206-12.
- Clark, M. G., "Tests of historical sense in fourth and fifth grade pupils," *Historical Outlook*, XIV (April, 1923) 147-150.
- Columbia Research Bureau American History Test*, Carman, H. J., Barrows, T. N., Wood, B. D. World Book Company, New York, 1926. Grades 10 to 14.
- Cooperative History Tests. Cooperative Test Service, New York City.
- Davis, S. B., *Exercises in United States History*. University of Pittsburg. Elementary grades.
- Davis, S. B., and Hicks, E. E., *Informational Test to Accompany Beard and Bagley's History of the American People*.
- DeGraff, M. H., "A measure of factual achievement in American history," *Historical Outlook*, XVI (November, 1925) 320-322.
- Denny, E. C., Nelson, M. J., *Denny-Nelson American History Test*. World Book Company, New York, 1928. Grades 7-8.
- Denver Public Schools, *Semester Test in World History I*, 1930; *Semester Test in World History II*, 1930. *Semester Test in American History*, 1930.
- Detroit Public Schools, Department of Research. *History 4 Examination*, May, 1926. *7B History Test, Form A, Character Judgment* (Selected from Pressey) *American History* (1) exam. f. B, by C. C. Barnes, E. M. Kimball, S. J. Palmer, P. T. Rankin and H. L. Simpson. *American History* (2) exam. f. B, Ditto. *World History* (1) Part I, by 12 authors, 1929. *World History* (1) Part II, by 12 authors, 1929. *History* (3), Parts I and II, by 12 authors, 1927. *History* (4), Parts I and II, by 10 authors, 1927.
- Dewey-Nelson American History Test*. World Book Company, New York. Grades 7-8.
- Diagnostic Test in Modern European History*. Indiana University Bookstore, Bloomington, Ind.

TESTS AND MEASUREMENTS

Elston, Bertha, "Improving the teaching of history in the high school through the use of tests," *Historical Outlook*, XIV (Nov., 1923) 300-05.

Ely-King Interpretation Tests in American History. Southern California School Book Depository, Hollywood, Calif. Grade 8. Seven unit tests as follows:

Test I. Periods of Exploration, Settlement, and Colonization.

Test II. The Struggle for Independence and the Founding of a New Nation.

Test III. Political and Territorial Growth.

Test IV. Development of the Nation.

Test V. The Slavery Problem, the Civil War, Reconstruction and the New South.

Test VI. Fifth Years of Progress, 1864-1914.

Test VII. Recent Events, Including the World War and Its Aftermath.

Fordyce Test for History VII (American) High School, Muskogee, Oklahoma.

Fordyce Test for History IV (Ancient) High School, Muskogee, Oklahoma.

Fordyce Test for History V (Medieval) High School, Muskogee, Oklahoma.

Fordyce Test for History VI (Modern) High School, Muskogee, Oklahoma.

Gibbons, A. N., *Factual Background Test in American History to 1787*.

Factual Background Test in Ancient and Medieval History.

Factual Background Test in Introduction to Modern European History.

History 1-2.

All published by National Council for the Social Studies, March, 1928.

Gibson, O. H., "Existing standard tests in history," *Historical Outlook*, XII (December, 1921) 324-326.

Gold, Mary, "Test of vocabulary in history," *Historical Outlook*, XVII (October, 1926) 285-291.

Gregory, C. A., *Tests in American History*. C. A. Gregory Co., Cincinnati, 1927. Grade 7; grade 8; and grades 8-12.

Gregory-Owens Test in Medieval and Modern History. C. A. Gregory Co., Cincinnati. High school and normal.

BIBLIOGRAPHY OF TESTS

- Griffith, G. L., "Harlan's American History Test in the New Trier Township Schools," *School Review*, XXVIII (November, 1920) 697-708.
- Haggerty, M. E. (See *National Intelligence Tests*).
- Hahn, H. H., Communication—"Standard Tests," *Historical Outlook*, XIII (Jan., 1922) 12.
- Hahn, H. H., *History Scale*. Public School Publishing Co., Bloomington, Ill. Grades 7-8.
- Harlan, C. L., "Educational measurement in the field of history," *Journal of Educational Research*, II (December, 1920) 849-853.
- Harlan, C. L., *Test of Information in American History*. Public School Publishing Company, Bloomington, Ill. Elementary grades.
- Harvard Comprehensive Exams.—History*.
- Hathaway, Gladys M., "Vocabulary difficulties in a 4th grade history test," *University of Pittsburgh School of Education Journal*, V (June, 1930).
- Hefley, Sue, "Test on world history," *Historical Outlook*, XX (January, 1929) 27-29.
- Henmon, V. A. C., "Some limitations of educational tests," *Journal of Educational Research*, VII (March, 1923) 185-198.
- Herrick Test of Generalizations*. Grades 11 and 12.
- Iowa General Information Tests in American History*, Bureau of Educational Research and Service, State University of Iowa, Iowa City. Junior and senior high school.
- Iowa Academic Contest**
- 1929: *Examination in American History: Test A.*
Examination in World History: Test A.
- 1930: *Examination in American History: Test A.*
Examination in World History: Test A.
- 1931: *Every-Pupil Test in United States History*, Anderson, H. R., and Andrews, G. G.
- 1933: *Every-Pupil Test in World History*, Anderson, H. R., and Andrews, G. G.
Every-Pupil Test in United States History, Anderson, H. R., and Andrews, G. G.
Every-Pupil Test in American Government, Anderson, H. R., and others.
Every-Pupil Test in Economics, Anderson, H. R., and others, edited by J. E. Partington.
Subject Matter Circular in World History, Anderson, H. R., Andrews, G. G. Edited by E. F. Lindquist.

TESTS AND MEASUREMENTS

- Junior American History Test.* World Book Company, New York. Junior high school.
- Kelley, T. L., *Test of Historical Ability*, Teachers College Contributions to Education, No. 71. High school grades.
- Kepner, Tyler, "An aspect of historical testing," *Historical Outlook*, XV (December, 1924) 414-416.
- Kepner, P. T., "A survey of the test movement in history," *Journal of Educational Research*, VII (April, 1923) 309-325.
- Knowlton, D. C., "Improving the quality of instruction in history with the aid of the photoplay. The development of teaching procedures and teaching techniques," *Historical Outlook*, XX (April, 1929) 167-179; (May, 1929) 229-230.
- Koch, Julia F., "An adaptation of the standardized test to medieval history," *Education*, L (December, 1929) 232-242.
- Krey, A. C., "What does the new-type examination measure in history?" *Historical Outlook*, XIX (April, 1928) 159-162.
- Krey, A. C., Wesley, E. B., "Does the new type test measure results of instruction in the social sciences? A report of an experiment," *Historical Outlook*, XXIII (January, 1932) 7-21.
- Lindquist, E. F., and Anderson, H. R., "Objectives testing in world history," *Historical Outlook*, XXI (March, 1930) 115-122.
- Lindquist, A. E., "The form of the American History Examination of the Cooperative Test Service," *Educational Record*, XII (Oct., 1931) 459-475.
- Lindquist, E. F., and Anderson, H. R., "The improvement of objective testing in history," *Second Yearbook of the National Council for the Social Studies*, pp. 97-117, 1932.
- Lindquist, E. F., and Anderson, H. R., "Achievement Tests in the Social Studies," *Educational Record* (April, 1933).
- Lynch, I. R., *Lynch American History Test Series*. Educational Test Bureau, Minneapolis, 1926. Grade 8.
- Martin, D. H., "A comparison of objective examinations in American history," *American School Board Journal*, LXXVI, no. 4 (April, 1928) 49, 130, 132.
- Maryland State Department of Education, *History Tests for Grades 4-7*.
- McCullum, D. F., *Test in American History*. *Journal of Educational Psychology*, Vol. VIII, 257-74.
- Metropolitan Achievement Tests*: New York City Edition—History. World Book Co., Yonkers. Grades 4-8.
- Michell, E., *Teaching Values in New Type History Tests*. World Book Co., Yonkers. 1930.

BIBLIOGRAPHY OF TESTS

- Minnesota State Board Examinations in Ancient and Modern History, May, 1928.
- Modern European History*. Harlow Publishing Co., Oklahoma City, Okla. High school.
- Modern European History Test*. Center for Psychological Service, Washington, D. C. High school or college.
- Modern School Achievement Tests—History and Civics*. Bureau of Publications, Teachers College. Grades 3-9.
- Monroe, Walter S., "Historical information and judgment in pupils of elementary schools," *Journal of Educational Research*, III (April, 1921) 303.
- Monroe, Walter S., "Written examinations and their improvement," *Historical Outlook*, XIV (November, 1923) 306-318.
- Moyer, F. E., "New types of history tests," *Historical Outlook*, XIV (November, 1923) 323-327.
- Myers, Carry C., "Delayed recall in history," *Journal of Educational Psychology*, VIII (May, 1917) 275-283.
- New Stanford Achievement Tests—History and Civics*. World Book Co., Yonkers. Grades 4-9.
- Odell, C. W., "The Barr Diagnostic Tests in American History," *School and Society*, XVI (October 28, 1922) 501-503.
- Penell, Olivia C., "Research tests in United States history," *Historical Outlook*, XV (March, 1924) 128-143.
- Peters, Charles C., and Altman, John E., "A critical study of the content of standardized tests in American History," *Journal of Educational Research*, XXIII (February, 1931) 153-161.
- Philadelphia Public Schools, Division of Educational Research, "Report on a survey test in history of the United States," *Historical Outlook*, XX (January, 1929) 17-23.
- Pratt, Orville C., "Spokane United States History Test," *Journal of Educational Research*, III (February, 1921) 155-157.
- Pressey, L. C., *Test of Special Vocabulary of History*. Public School Publishing Company, Bloomington, Ill.
- Pressey, L. C., Richards, R. C., *Understanding of American History*. Public School Publishing Company, Bloomington, Ill. Grades 6-12.
- Pressey, S. L., "An examination for measurement in history, arithmetic and English," *Journal of Educational Research*, III (May, 1921) 359-69.
- Public School Achievement Test—History*. Public School Publishing Co., Bloomington, Ill., 1928. Grades 5-8.
- Ralston, Sue W., "The use of supplementary readings and fact tests in connection with a one year course in ancient and medieval history," *Historical Outlook*, XV (December, 1924) 410-414.

TESTS AND MEASUREMENTS

- Rayner, *American History Test*. University of Illinois Bureau of Research. Elementary grades.
- Renfrow *Tests in European Background to American History*. C. A. Gregory Co., Cincinnati. Grade 6.
- Ruch, G. M., *General Information Tests in American History*. Commonwealth investigation exam's soc. sci. University of Iowa.
- Rugg, Earle U., "Character and value of standardized tests in history," *School Review*, XXVII (December, 1919) 757-772; or *Historical Outlook*, XI (June, 1920) 231-236.
- Sackett, Leroy W., "A scale in ancient history," *Journal of Educational Psychology*, VIII (May, 1917) 284-293.
- Sackett, Leroy W., "A scale in United States history," *Journal of Educational Psychology*, X (September, 1919) 345-348.
- Shryock, Richard H., "New tests for old. An examination of the relationship of aims and methods of testing in the teaching of history," *Historical Outlook*, XIV (November, 1923) 319-322.
- Sloyer *Test in World History*. Palmer Co., 120 Boylston St., Boston.
- Spokane History Test* (O. C. Pratt, Superintendent).
- Starch, D., *American History Tests*. University Cooperative Co., Madison, Wis.
- Stormzand, M. J., "Tests in American history," *American History Teaching and Testing*. Macmillan, New York, 1925. 50ff.
- Thornton, E. W., "The use of informational tests in American history teaching," *Historical Outlook*, XX (January, 1929) 12-16.
- Tyrrell *American History Exercises*, 120 Boylston St., Boston.
- United States History*. Harlow Publishing Co., Oklahoma City, Okla. Grades 7-8.
- Unit Scales of Attainment—American History, United States History*. Educational Test Bureau, Minneapolis. American, grades 3-6; United States, grades 7-8.
- Van Bibber, L. C., "Study of specific classroom difficulties in teaching history," *Second Yearbook*, National Council for Social Studies, 1932.
- Vannest, C. G., *Diagnostic Test in Modern European History*. Indiana University Bookstore, Bloomington, Ind. High school grades.
- Van Wagenen, M. J., *American History Scales*. Bureau of Publications, Teachers College, 1919. Grades 5-6.
- Van Wagenen, M. J., "The Van Wagenen reading scales in history, general science, and English literature," *Journal of Educational Research*, III (April, 1921) 314-316.

BIBLIOGRAPHY OF TESTS

- Weaver, Robert B., and Traler, Arthur E., "Essay examinations and objective tests in United States history in the junior high school," *School Review*, XXXIX (November, 1931) 689-695.
- Wessel, Herman M., "A critical study of the results of Barr's Diagnostic Tests in American History," *Historical Outlook*, XVII (May, 1926) 230-31.
- Williams, L. W., "Barr's Diagnostic Tests in American History," *Journal of Educational Research*, III (January, 1921) 74-75.
- Wilson, H. C., and Waples, D., *Class room procedure test in history and other social studies*. University of Chicago Press, 1928.
- Wilson, H. E., "The continuity test in history teaching," *School Review*, XXXIV (November, 1926) 679-684.
- Witham, E. C., *Comprehensive History Tests*, J. L. Hammett Company, Cambridge, Mass. Grades 7 and 8.
- Wood, B. D., *The I. E. R. College Entrance Examinations in Ancient History*. Teachers College, Bureau of Educational Research, 1922.

Civics and Government

- Abelow, S. P., "A group test of civics ability," *Historical Outlook*, XXII (November, 1931) 345-349.
- Almack Tests in American Civics and Government*. C. A. Gregory Co., Cincinnati. High School.
- American Civics and Government Test for High Schools*. Public School Publishing Co., Bloomington, Ill.
- American Council Civics and Government Test*—Leigh, R. D., McGoldrick, J. D., Odegard, P. H., Wood, B. D., World Book Company, 1929. High schools and colleges.
- American Government and Civics*. Harlow Publishing Co., Oklahoma City, Okla.
- Barr, A. S., "Measurements in civics," *Historical Outlook*, XV (January, 1926) 24-26.
- Bear Test on the United States Constitution*. Public School Publishing Co., Bloomington, Ill.
- Brown, A. W., and Woody, C., *Brown-Woody Civics Test*. World Book Co., 1928.
- Burton, W. H., and V. N., *Civics Test*. World Book Company, 1928.
- Cannon, Annie E., "A plan for teaching a course in civics on the unit basis," *Historical Outlook*, XVI (January, 1925) 28-35.
- Chassell-Upton Citizenship Scales*. Teachers College Bureau of Publications, 1922.
- Civics*. Harlow Publishing Co., Oklahoma City, Okla.

TESTS AND MEASUREMENTS

Dawson, Edgar, "The history inquiry . . . VII. Experimenting with a test," *Historical Outlook*, XV (June, 1924) 260-268.

Gibbons, A. N., *Civics Final Test.*]

Background Test in Civics

Harvard Comprehensive Exams.—Government. All published by National Council for the Social Studies, March, 1928.

Politics

Hill, H. C., *A Test in Civic Information.* Public School Publishing Company, 1926. Grades 6-12.

Hill, H. C., and Wilson, H. E., *Test in Civic Action.* Public School Publishing Company, Bloomington, Ill., 1929. Grades 6-12.

Iowa Academic Contest, 1930: *Examination in American Government.*

Johnstown Public Schools, *Community Civics Test.* Johnstown, Pa.
Minneapolis Public Schools Curriculum Department, *Test in Understanding of Ideas Used in Community Life Problems Course.*

Monroe, Walter S., "Written examinations and their improvement," *Historical Outlook*, XIV (November, 1925) 306-318.

New York State Civics Test. State Education Department, Albany, 1928.

Rugg, H. O., *Vocational Civics*, Grade 7B-7A. Based on *An Introduction to American Civilization.* Ginn & Co., 1929. Test by Miss Tulley, Providence, R. I.

Stack, H. J., "Standardized tests in community and economic civics," *Historical Outlook*, XVIII (April, 1927) 166-172.

Stephenson, O. W., and McGehre, W. R., "A Vocabulary common to citizenship and American history," *Journal of Educational Research*, XXII (June, 1930) 55.

Upton-Chassell Scale for Measuring Habits of Good Citizenship. Teachers College Bureau of Publications, 1922.

Economics

American Council Economics Test (H. Taylor, T. N. Barrows, B. D. Wood) World Book Company, 1929.

Iowa Academic Contest, 1930: *Objective Examination in Principles of Economics.*

Kornhauser, A. W., "Results from the use of objective tests and questionnaires in The Economic Order," *Journal of Business of the University of Chicago*, I (October, 1928) 429-455; II (January, 1929) 82-111.

Sins, V. M., *Score Card for Socio-Economic Status.* Public School Publishing Co., Bloomington, Ill., 1927.

BIBLIOGRAPHY OF TESTS

Geography

- Analytical Scales of Attainment in Geography.* Educational Test Bureau, Minneapolis.
- Barthelmess, H. M., "Geography testing in Boston." *Journal of Educational Research*, II (November, 1920) 701-712.
- Branom, M. E., *Diagnostic Tests in Geography.* McNight and McNight, Normal, Ill., 1926. Grades 5-8.
- Buckingham, B. R., and Stevenson, P. R., *Geography Test*, United States. Information-Problems. Public School Publishing Co., Bloomington, Ill.
- Buckingham, B. R., Stevenson, P. R., *Place Geography Tests.* Public School Publishing Co., Bloomington, Ill., 1922.
- Cavins, L. V., "An experiment with standardized tests in a state teachers examination." *Journal of Educational Research*, XIV (October, 1926) 206-12.
- Courtis, S. A., *Standard Test in Geography.* Detroit, Mich. Elementary grades.
- Detroit Public Schools, Department of Research
Geography Test 1 b S—The world, oceans and continents
Geography Test 2 b S—Location, states
Geography Test 3 b N—Location, countries
Geography Test 4 b S—Location, countries of Europe
Geography Test 5 b O—Location, countries and cities of western hemisphere.
- Forney, E. B., *Test in Map-Reading Abilities.* Ginn & Co., 1926.
- Geography Test.* Harlow Publishing Co., Oklahoma City, Okla. Grades 7-8.
- Gregory, C. A., Spencer, P. L., *Geography Tests.* University of Cincinnati, 1923. Grades 6-8.
- Geography Tests.* Plymouth Press, Chicago, Ill. Elementary grades.
- Gregory-Haggerty Geography Tests.* C. A. Gregory Co., Cincinnati. Grades 4-6.
- Hahn, H. H., Lackey, E. E., *Geography Scale.* Public School Publishing Co., 1928. Elementary grades.
- Hartwell, *Geography Test.* Mentzer, Bush and Co., Chicago.
- Industrial Commercial Geography.* Harlow Publishing Co., Oklahoma City, Okla. High school.
- Information-Problems Test in Geography.* Public School Publishing Co., Bloomington, Ill. Elementary grades.
- Johnstown (Pa.) Public Schools, *Geography Test.*

TESTS AND MEASUREMENTS

- Lackey, E. E., "A scale for measuring the ability of children in geography," *Journal of Educational Psychology*, II (October, 1918) 443-451.
- Mathewson, Chester A., "A preliminary note on the use of the Hahn-Lackey Geography Scale," *Journal of Educational Psychology*, IX (October, 1918) 467.
- Mathewson, Chester A., "Source results with the Hahn-Lackey Scale in Geography," *Journal of Educational Psychology*, II (December, 1918) 581-87.
- Metropolitan Achievement Tests: New York City Edition—Geography.* World Book Co., Yonkers.
- Modern School Achievement Tests—Geography.* Bureau of Publications, Teachers College.
- New Stanford Geography Test.* World Book Co., Yonkers. Grades 4-9.
- Olmstead, M. C., *Diagnostic Geography Test.* Junior High School Clarkston, Washington. Elementary grades.
- Place Geography Tests.* Public School Publishing Co., Bloomington, Ill.
- Posey, C. J., Van Wagenen, M. J., *Posey-Van Wagenen Geography Scales.* Public School Publishing Company, Bloomington, Ill., 1922. Grades 5-8.
- Public School Achievement Test—Geography.* J. S. Orleans, Public School Publishing Co., Bloomington, Ill., 1928. Grades 6-8.
- Russell-Harr Unit Tests in Geography.* Bureau of Educational Measurements and Standards, Kansas State Teachers College, Emporia, Kansas. Grades 4-8.
- Stevenson, P. R., "A problem test in geography (a preliminary statement)," *Journal of Educational Research*, V (April, 1922) 350-353.
- Stevenson, P. R., "Minimum essentials in place geography," *Journal of Educational Research*, VI (November, 1922) 300-306.
- Unit Scale of Attainment—Geography.* Educational Test Bureau, Minneapolis.
- Wiedefeld-Walther Geography Test.* World Book Co., Yonkers. Grades 4-8.
- Witham Geography Test of Europe.* Southern California School Book Depository, Hollywood, Calif. Grades 6-9.
- Witham Geography Test of the United States.* Southern California School Book Depository, Hollywood, Calif. Grades 5-8.

BIBLIOGRAPHY OF TESTS

Witham Standard Geography Tests. J. L. Hammett Co., Cambridge, Mass. Grades 5-12.

Witham Tests in World Geography. Southern California School Book Depository, Hollywood, Calif.

The Social Studies

(general field of)

Allport, F. H., and G. W., "Personality Traits," *Journal of Abnormal and Social Psychology*, 1921.

Allport, W., and Vernon, P. F., *A Study of Values*. Houghton Mifflin, 1931.

Alpha Group Examination—U. S. S., Medical Dept., Division of Psychology, 1918.

Association Press—*Bible Knowledge Test, Series of Character and Personality Tests*. New York, 347 Madison Ave.

Association Press—*Character Growth Tests, Christian World Citizenship*, 1926-27.

Association Press—*Character Growth Tests, Group Relationships*, 1926-27.

Association Press—*Character Growth Tests, Immediate Life Situations*, 1926-27.

Association Press—*Character Growth Tests—Religious Concepts*, 1926-27.

Association Press—*General questions for Boys*, 1927-28.

Association Press—*Health Knowledge Test*.

Association Press—*Interest Analysis*.

Association Press—*Laxcock Test of Biblical Information*.

Association Press—Lehman, H. C., *Lehman Play Quiz*.

Association Press—Watson, G. B., *Some Methods of Program Appraisal*, 1926.

Association Press—Watson, G. B., *Rating Scales*, 1928.

Association Press—Watson, G. B., *Experiments with Religious Education Tests*, 1926.

Association Press—Watson, G. B., *Some Accomplishments in Summer Camps*, 1926.

Association Press—*Test of Religious Thinking, Forms A, E*.

Atlanta Public Schools, Department of Testing and Guidance, *Atlanta Social Science Examination*.

Anonymous, "Finding list of hopeful undertakings," *Historical Outlook*, XIII (December, 1922) 321-326.

Baker, Harry J., *Telling What I Do Test*. Public School Publishing Co., Bloomington, Ill., 1930.

TESTS AND MEASUREMENTS

- Berne, Esther Van Cleave, *An Experimental Investigation of Social Behavior Patterns in Young Children*. 1928. Unpub. Ph.D. Thesis. University of Iowa.
- Berne, Esther Van Cleave, *An Investigation of the Wants of Children*. Iowa Child Welfare Research Station, 1929.
- Bibliography of Objective Tests in the Social Sciences*. University of the State of New York.
- Bowman, K. M., *Mental Disease Research Questionnaire*. (Revised June 15, 1929). Boston Psychopathic Hospital.
- Bureau of Personnel Research, Downey, J. E., *Test II*. 1920.
- Bureau of Personnel Research, *Test XI*. 1920.
- Burton Social Science Civics Tests*, C. A. Gregory Co., Cincinnati. Grades 7-12.
- Character Education Institution, *The Characteristics of Human Beings*. National Capital Press, Washington, D. C.
- Chassell, J. O., *Experience Variables*. University of Rochester Medical School, 1928.
- Columbia Course in Contemporary Civilization*, Test in Midyear, 1925.
- Cornell, E. L., and Coxe, W. W., Orleans, J. S., *Rating Scale for School Habits*. World Book Company, Yonkers, 1927.
- Cowdery, Karl M., *Interest Report Blank*. Revision of C. I. T. Bureau of Personnel Research Interest Analysis Blank.
- Denver Public Schools, *Curriculum Tests in Social Science*, Grades 7B, 7A, 8B, 8A, 9B, 9A.
- Detroit Public Schools, *Character Education in Detroit*, Report of Committee on Character Education. 1927.
- Detroit Public Schools, Department of Research
Social Science Test VI, form A—Europe, by M. E. Irwin and L. Spencer
Social Science Test VII, form A—U. S., by M. E. Irwin and L. Spencer
Social Science Test VIII, form A—Grade 5A by M. E. Irwin and L. Spencer
Social Science Test X, Part I—Foods. Grade 3b, by 5 authors.
 Part II—Shelter and Fuel. Grade 3B, 5 authors.
 Part III—Clothing. Grade 3B, 5 authors.
Social Science Vocabulary.
- Detroit Teachers College, *How Children Choose Friends*. 1929.
- Dimock, H. S., Hendry, C. E., *Camping and Character*. Association Press, 1929, New York.

BIBLIOGRAPHY OF TESTS

- Doll, E. A., *Vineland Adjustment Score Card*. Publication of the Training School at Vineland, N. J.
- Downey, J. E., *Individual Will-Temperament Test*. World Book Co., 1921.
- Floyd, Oliver R., "Validity of tests in the social sciences," *Historical Outlook*, XX (January, 1929) 10-12.
- Franzen, R., *Health Education Tests*. American Child Health Association, New York. Grades 5-6.
- Furfey, P. H., "Developmental Age," *American Journal of Psychology*, VIII (July, 1928). Revised 1930. C. H. Stoelting Co., Chicago.
- Furfey, P. H., *The Gang Age*. New York, Macmillan, 1926.
- Furfey, P. H., "Some factors influencing the selection of boys' chums," *Journal of Applied Psychology*, 1927.
- George Washington University Series *Social Intelligence Test*, Moss, F. S., Hunt, T., Omwake, K. T., Ronning, M. M. 1927
- Gibbons, A. N. *Tests in the Social Studies*.
Social Science Reading Composition.
Working Skill Test in Social Science Material.
Published by National Council for the Social Studies, No. 3, Athens Press, Iowa City, 1928.
- Haggerty, M. E., Olson, W. C., Wickman, E. K., *Behavior Rating Schedules*. World Book Co., 1930.
- ✓Hardy, Ruth, "New Types of tests in social science," *Historical Outlook*, XIV (November, 1923) 326-328.
- Harvard Background Tests in Social Studies*. Ginn and Co., Boston.
- Heidbreder, E., *Personality Traits Rating Scale*. Univ. of Minn.
- Hill, H. C., "The use of tests in the teaching of the social studies," *Historical Outlook*, XX (January, 1929) 7-10.
- House, S. Daniel, "A Mental hygiene inventory," *Arch. of Psychology*, No. 88 (May, 1927). (Use with Woodworth Test.)
- Iowa High School Content Examination*. Bureau of Educational Research and Service, State University of Iowa.
- Kelley, T. L., *Scientific Method*. MacMillan, 1932.
- Kelley, T. L., "The objective measurement of the outcomes of the social studies," *Historical Outlook*, XXI (Feb., 1930) 66-72.
- Kelley, T. L., *Values in High School Algebra, and their Measurement*, Teachers College Bureau of Publications. May, 1920.
- Kelley, T. L., Ruch, G. M., and Terman, L. M., *Stanford Achievement Test*. Grades 4-10. 1 of 10 sections is geography, 1 of 10 sections is history information.

TESTS AND MEASUREMENTS

- Kelty, M. G., and Moore, N. E., *Test of Concepts in the Social Studies*. Scribners, New York, 1934.
- Kepner, T., *Harvard Tests, Background in Social Sciences*. Ginn and Co., 1924. High school grades.
- Kimmel, W. G., "The use of practice tests in the teaching of the social sciences," *Historical Outlook*, XIV (December, 1923) 354-358.
- Kohs, S. C., *Ethical Discrimination Test*. C. H. Stoelting Co., Chicago, 1922.
- Kregel, L. J., "A summary of social science tests for the junior high school," *Journal of Educational Research*, XX (November, 1929) 308-17.
- Krey, A. C., and Wesley, E. B., "Does the new type test measure results of instruction in the social studies?" *Historical Outlook*, XXIII, Jan., 1932.
- Lewerenz-Steinmetz *Orientation Test*. Southern California School Book Depository, Hollywood, California.
- Lincoln School: *Town and City Life in America—Resources, Industries and Cities in America*, Rugg, H. O., Earle R., Schweppe, E. *Social Science Foundations Test* (See vol. 1) Rugg, H. O., Showalter, D. R., Mathews, C. O.
- Maller, J. B., *Character and Personality Tests* (Bibliography). Bureau of Publications, Teachers College, 1930.
- Maller, J. B., *Self-Marking Test* (Objective Test of Honesty). Bureau of Publications, Teachers College, 1930.
- Marshall, L. C., *Request for Aid in Formulating a Statement of Basic Processes in Society*. University of Chicago.
- Martz, C. E., "Old and new diagnosis and remedies in senior high school social studies," *Historical Outlook*, XX (May, 1929) 216-218.
- Mathews, C. O., *Tests on the Social Studies*. Bureau of Publications, Teachers College.
- May, M. A., *Reason Why You Go to Church Scale*. Teachers College, Columbia University.
- McDonough, Sister Mary Rosa, *An Empirical Study of Character*. The Catholic University of America, Washington, D. C., 1929.
- Meltzen, H., *Children's Social Concepts*. Teachers College Contributions to Education, No. 192. 1925.
- Minnesota University—Publications of the Institute of Child Welfare.

BIBLIOGRAPHY OF TESTS

- Moore, E. S., *Development of Mental Hygiene in Young Children through Purposeful Activity*. University of Iowa Child Welfare Institute, 1929.
- National Intelligence Tests*, National Research Council. Haggerty, M. E., Terman, L. M., Thorndike, E. L., Whipple, G. M., Yerkes, R. M. World Book Co., 1920.
- New York Rating Scale for School Habits*, Cornell, E. L., Cox, W. W., and Orleans, J. S.
- Orleans, J. S., *Public School Achievement Tests*. Public School Publishing Co., Bloomington, Ill., 1928.
- Pennsylvania College Press and State Department of Public Instruction, *College Achievement Test*—Carnegie Foundation, New York, 1928.
- Pressey, L. C., *Technical Vocabularies of Public School Subjects*, Public School Publishing Co., Bloomington, Ill.
- Pressey, L. C., *X-O Tests*. C. H. Stoelting Co., Chicago.
- Public School Achievement Test—J. Health*. Orleans, J. S. Grades 4-8. Public School Publishing Co., Bloomington, Ill.
- Religious Education Association, Second Annual Religious Conference, 1929.
- Rugg, Hockett, and others, *Tests on the Rugg Social Science Pamphlets*. Bureau of Publications, Teachers College, Columbia University.
- Sangren, P. V., "The present status of measurement in the social sciences," *Historical Outlook*, XXI (October, 1930) 279-283.
- Schwesinger, G. C., *Social-Ethical Significance of Vocabulary*, Teachers College, Contributions to Education, No. 211, 1926.
- Showalter, D. R., *Social Studies Inventory Tests*, in cooperation with The Social Science Pamphlet Investigation, Lincoln School of Teachers College, New York City.
- Sones, W. W. D., Harry, D. P., Jr., *Sones-Harry High School Achievement Test*. World Book Company, 1929. Grades 9-13.
- Strong, E. K., Jr., *Vocational Interest Blank*. Stanford University, 1929.
- Symonds, P. M., *List of Acts which Indicate Absence of Tact or Social Adjustment*. Teachers College, Columbia University.
- Thurstone, L. L., *Employment Tests*. World Book Company, 1922.
- Thurstone, L. L., "Fechern's Law and the method of equal-appearing intervals," *Journal of Experimental Psychology*, June, 1929.
- Tomlin, F. E., *Best Thing to Do—A Test of Knowledge of Social Standards*. Stanford University Press, 1931.

TESTS AND MEASUREMENTS

- Trabue, M. R., *North Carolina High School Examination*. Bureau of Educational Research, University of North Carolina, 1929.
- Tryon, R. M., "Standard and new type tests in the social studies," *Historical Outlook*, XVIII (April, 1927) 172-78.
- Waples, D., *Comparative Study of Adult Reading Interests*, University of Chicago, 1928-1929.
- Waples, D., Teachers' Reading of Non-Fiction, *Educational Research Bulletin*, Vol. VIII, No. 17, Nov., 1929.
- Watson, G. B., *Analytical Score Sheets for Watson Test of Public Opinion*, Teachers College Bureau of Publications, 1927.
- Watson, G. B., "A scale for rating home contribution to personality development of children," *Baltimore Bulletin of Education*, Vol. 8, No. 8, (May, 1930).
- Wesley, E. B., *Tests in Social and Political Terms*. Charles Scribner's Sons, 1932.
- Wyman, J. B., Interest Tests. Dissertation in files of Stanford University Library.

Social Attitudes

- Allport, G. W., "Composition of political attitudes," (bibliog.) *American Journal of Sociology*, XXXV (Sept., 1929) 220-28.
- Andrews, J. M., "Development of wholesome attitudes," *National Educational Association Journal*, XVIII (December, 1929) 305-6.
- Association Press—*A Study of Opinions, Feelings, and Attitudes Concerning Some International Problems*. New York, High school and college.
- Association Press—*Character Growth Tests, Personal Attitudes*. New York, 1926-27.
- Association Press—*Measurement of Personal Attitudes in Younger Boys*, Sweet, L. 1929, New York.
- Association Press—*Opinion on International Questions*, Neuman, G. B., Hollinger, R. W., Porter, E., Frederick, R., Keeny, S. M. New York.
- Association Press.—*Opinions on Race Relations*. New York.
- A Test of Social Beliefs and Attitudes*. Bureau of Publications, Teachers College.
- Bain, Read, "An Attitude on Attitude Research," *American Journal of Sociology*, XXXIII (1928) 941-57.
- Bain, Read, "Religious attitudes of college students," *American Journal of Sociology*, XXXII (1927) 762-70.

BIBLIOGRAPHY OF TESTS

- Betts, G. H., "Religious attitudes and activities of university students," *Religious Education*, XXIII (1928) 917-19.
- Biddle, William W., "The relationship between knowledge and a measure of autistic thinking on certain international problems," *Journal of Social Psychology*, 1931, 493.
- Bogardus, E. S., "Social distance and its origins," *Journal of Applied Psychology*, IX (1925) 215-26.
- Bogardus, E. S., *Leadership and Attitudes. Bibliography of Sociology and Social Research*, XIII (March, 1929) 377-81.
- Book, W. F., "On the genesis and development of conscious attitudes," *Psychological Review*, XVII (1910) 381-398.
- Briggs, T. H., "Enlarged Concept of Secondary Education," Kentucky University *Proceedings of Seventh Annual Education Conference.*, 19-28.
- Butler, G., "Building correct attitudes," *Agricultural Education*, I (March, 1929) 12.
- Campbell, H. M., "Prejudiced thinking of school children," *Teachers Journal and Abstract*, V (October, 1930) 443.
- Chambers, M. M., "Inculcating respect for law in the schools," *Education*, L (February, 1932) 356-58.
- Clark, W. W., "Measurement of social attitudes," *Journal of Applied Sociology*, VIII (1924) 345-54.
- Clave, E. J., "New type scale of measuring attitude," *Religious Education*, XXIII (April, 1928) 364-68.
- Davis, Jerome, "Testing the social attitudes of children in the government schools of Russia," *American Journal of Sociology*, XXXII (1927) 271-81, 947ff.
- Droba, D., *Measurement of Attitude Toward Militarism, Pacifism.* Unpub. (bibliog. of Hinckley below --University of Florida)
- Dudycha, D. J., "The religious beliefs of college freshmen," *School and Society*, XXXI (1930) 206-8.
- Dudycha, D. J., "Social beliefs of college freshmen," *School and Society*, XXXII (December 20, 1930) 846-9.
- Everett, Samuel, "Objective tests the best discoverer of pupil attitudes," *Historical Outlook*, XX (November, 1929) 335-337.
- Farnsworth, Paul, and Behmer, Alice, "A note on the attitude of social conformity," *Journal of Social Psychology*, 1931, 126.
- Frederick, Robert, "An investigation into some social attitudes of high school pupils," *School and Society*, XXV (1927) 410ff.
- Freeman, G. L., "Influence of attitude on learning," *Journal of Genetic Psychology*, III (January, 1930) 98-111.
- Garrison, K. C., and Mann, Margaret, "A study of the opinions of college students," *Journal of Social Psychology*, 1931, 168ff.

TESTS AND MEASUREMENTS

- Gates, Georgiana S., "An experimental study of the growth of social perception," *Journal of Educational Psychology*, XIV (1923).
- Glen, M., "Attitudes toward music," *Music Supervisors Journal*, XV (March, 1929) 75-84.
- Guilford, J. P., "Racial preferences of a thousand American university students," *Journal of Social Psychology*, II (1931) 179ff.
- Hamilton, C. H., *Religious Education and Attitude Test*, Form B. Virginia Agricultural Experiment Station and Department of Agriculture, U. S. A., publication.
- Harper, M. H., *Social Beliefs and Attitudes of American Educators*. Teachers College, Contributions to Education, No. 294, 1927.
- Harper, H. R., *Study of Opinions, Feelings and Attitudes Concerning Some International Problems*, Teachers College, 1929.
- Harper, H. R., *What European and American Students Think on International Problems*, Teachers College, 1931.
- Hart, Hornell N., *A Test of Social Attitudes and Interests*. University of Iowa, 1923.
- Herriot, M. E., *Attitudes as Factors of College Success*. bibliog. University of Illinois, 1929.
- Hinckley, E. D., "The influence of individual opinion on construction of an attitude scale," *Journal of Social Psychology*, August, 1932.
- Jones, E. S., "Opinions of college students," *Journal of Applied Psychology*, IV (1926) 427-36.
- Katz, Daniel, *Students' attitudes*, Syracuse, New York, 1931.
- Katz, D., and Allport, F. H., *Students' Attitudes; a Report of the Syracuse Reaction Study*. Craftman's Press, Syracuse, N. Y., 1931.
- Kornhauser, A. W., "Changes in the information and attitudes of students in an economics course," *Journal of Educational Research*, XX (November, 1930) 288-98.
- Kulp, D. H., and Davidson, H. H., "Can Neumann's Attitude Indicator be used as a test?" *Teachers College Record*, XXXII (January, 1931) 332-37. Bibliography.
- Lapierre, R. T., "Race prejudice: France and England," *Social Forces*, VII (1928) 102ff.
- Lasker, Bruno, *Race Attitudes in Children*. Holt, New York, 1927, 1929.
- Lasker, Bruno, "Childhood prejudices," *Child Study*, VI (February, 1929) 107-9.
- Lockhart, E. G., *Attitudes of Children toward Law*. University of Iowa, 1930.

BIBLIOGRAPHY OF TESTS

- Millson, W. A. D., "Experimental work in audience reaction," *Quarterly Journal of Speech*, XVIII (Feb., 1932) 13-20.
- Neumann, G. B., *A Study of International Attitudes of High School Students*. Bureau of Publications, Teachers College, 1926.
- Neumann, G. B., Kulp, D. N., and Davidson, Helen, *Test of International Attitudes*. Bureau of Publications, Teachers College, 1926, 1931.
- Oleson, D. M., and Jones, V., "Objective measure of emotionally toned attitudes," *Pedagogical Seminar*, XXXIX (June, 1931) 174-96.
- Peterson, R. C., and Thurstone, L. L., "Effect of a motion picture film on children's attitudes toward Germans," *Journal of Educational Psychology*, XXIII (April, 1932) 241-46.
- Powers, Francis, "The influence of intelligence and personality traits upon false beliefs," *Journal of Social Psychology*, (1931) 490-93.
- Riciardi, N., "What is young America thinking?" *Sierra Educational News*, XXV (November, 1929) 49.
- Rosander, A. C., *Can Attitudes Be Measured?* Phi Delta Kappan, October, 1930.
- Scales for Measuring Attitude toward War, the Church, Prohibition, etc.*, University of Chicago Press.
- Smith, H. L., *Measurement of Attitudes toward Prohibition*. Unpubl. University of Florida.
- Snyder, M. A., "Comparison of mental traits and attitudes of delinquent boys and girls," *Journal of Juvenile Research*, XV (July, 1931) 181-91.
- Stanford, A. T., "Study in social attitudes of a group of high school boys and girls," *School and Society*, XXVI (Dec. 3, 1927) 723-26.
- Stowe, E. M., "Student attitudes at Fukien Christian University," *Educational Review of China*, XXI (July, 1929) 287-93.
- Strebel, R. F., and Morehart, G. C., "Attitudes." *Nature* 66-69.
- Sullivan, Ellen B., *Attitude in Relation to Learning*. Princeton, and Albany, N. Y., Psychological Review Co., 1927.
- Sweet, Lennig, *The Measurement of Personal Attitudes in Younger Boys*. Association Press, New York, 1929.
- Thurstone, L. L., *The Measurement of Attitude*, University of Chicago Press, 1929.
- Thurstone, L. L., and Clave, E. J., *Measurement of Attitude toward the Church*. Unpubl.
- Thurstone, L. L., "The Measurement of Opinion," *Journal of Abnormal Psychology*, XXII (1928) 415-30.

TESTS AND MEASUREMENTS

- Thurstone, L. L., "An Experimental Study of nationality preference," *Journal of Genetic Psychology*, I, (1928) 405-25.
- Thurstone, L. L., "Theory of attitude measurement," *Psychological Review*, XXXVI (May, 1929) 222-41.
- Thurstone, L. L., "Attitudes can be measured," *American Journal of Sociology*, XXXIII (Jan., 1928) 533.
- Thurstone, L. L., "Influence of motion pictures on children's attitudes," *Journal of Social Psychology*, (1931) 291.
- Thurstone, L. L., "Measurement of social attitudes," Institute for Administrative Officers of Higher Institutions, *Proceedings*, 1931, III, 239-49.
- Tibbetts, F. L., "Student attitudes in junior college," *Junior College Journal*, II (December, 1931) 148-52.
- Vetter, G. B., "Measurement of social and political attitudes and the related personal factors," *Journal of Abnormal Psychology*, XXV (July, 1930) 149-81. Bibliog.
- Watson, G. B., "Does world-mindedness depend on goodwill or information upon character intelligence?" *Religious Education*, XXI (1926) 88-94.
- Watson, G. B., *Experimentation and Measurement in Religious Education*. Association Press, 1921, New York.
- Watson, G. B., "Orient and Occident, an opinion study," *Religious Education*, XXIV (1924) 322-328.
- Watson *Test of Public Opinion*. Bureau of Publications, Teachers College.
- Watson, G. B., *The Measurement of Fairmindedness*. Bureau of Publications, Teachers College.
- Willey, M. M., "Identification and inculcation of social values," *Bibl. of Ann. Am. Acad.*, CLX (May, 1932) 103-9.
- Willoughby, Raymond R., "A sampling of student opinion," *Journal of Social Psychology*, (1930) 164.
- Young, Kimball, *Measurement of Personal and Social Traits*. Publications of American Sociological Society, XXI (1927) 92-105.
- Young, Kimball, (editor) *Social Attitudes*. Holt, New York, 1931.

Test Construction

- Anonymous, "The construction of tests in the short-answer form," *Public Personnel Studies*, VII (1927) 130-37.
- Anonymous, "The statistical evaluation of test results," *Public Personnel Studies*, VI (1928) 2-11.
- Bayles, E. E., "Testing for comprehension," *Journal of Educational Research*, XX (1929) 268-272.

BIBLIOGRAPHY OF TESTS

- Brinkmeyer, I. H., "Sentence length as a determiner in true-false statements," *Journal of Educational Research*, XXII (1930) 203-204.
- Brinkmeyer, I. H., and Ruch, G. M., "Specific determiners in true-false statements," *Journal of Educational Research*, XXII (1930) 110-118.
- Buckingham, B. R., "A brief method of evaluating test material," *Journal of Educational Research*, XIX (1929) 159-168.
- Clark, L. E., "A method of evaluating the units of a test," *Journal of Educational Psychology*, XIX (1928) 263-265.
- Cleeton, G. U., "Optimum difficulty of group test items," *Journal of Applied Psychology*, X (1926) 327-340.
- Corey, S. M., "The effect of weighting exercises in a new type of examination," *Journal of Educational Psychology*, XXI (1930) 383-385.
- Curtis, F. D., and Woods, G. H., "A study of modified form of the multiple response test," *Journal of Educational Research*, XVIII (1928) 211-219.
- Estabrooks, G. H., "A new type of objective examination," *Pedagogical Seminary*, XXXIV (1927) 368-372.
- Giles, T. J., "Improving the objective test question," *School Review*, XXXV (1927) 286-88.
- Greene, Harry A., "Educational tests and the classroom teacher," *Extension Bulletin*, Iowa City, Iowa. College of Education, Series No. 25, 1928.
- Greene, Harry A., and Jorgenson, A. N., *The Use and Interpretation of Educational Tests*. Longmans, Green and Co., New York, 1929.
- Holmes, G. N., *A Statistical Study of a New Type of Objective Examination*. M.A. Thesis, University of Minnesota, 1930.
- James, H. W., "Technique in the construction of teachers' own objective tests," *Peabody Journal of Education*, IV (1927) 240-243.
- Jones, Edward Safford, *Comprehensive Examinations in American Colleges*. Macmillan, New York, 1933.
- Kelley, T. L., *The Interpretation of Educational Measurements*. World Book Company, Yonkers, 1927.
- McClusky, H. Y., and Curtis, F. D., "A modified form of the true-false test," *Journal of Educational Research*, XIV (1926) 213-224.

TESTS AND MEASUREMENTS

- Miller, G. F., "A variation in the true-false test," *School and Society*, XX (1924) 250-51.
- Opdyke, J. B., "Constructive Examinations," *Educational Review*, LXXIII (1927) 33-43.
- Orleans, J. S., and Sealey, G. A., *Objective Tests*. World Book Company, 1928.
- Paterson, D. G., "Improving the examining function in teaching," *Bulletin of the University of Minnesota*, XXVI (1923) 47-56.
- * Paterson, D. G., *Preparation and Use of the New-type Examination*. World Book Company, Yonkers, 1926.
- Peters, C. C., "The relation of standardized tests to educational objectives," *Second Yearbook of the National Society for the Study of Educational Sociology*, (1929) 148-159.
- Ruch, G. M., *The Objective or New-type Examination*. Scott-Foresman, 1929.
- Ruch, G. M., and Rice, G. A., *Specimen Objective Examinations*. Scott-Foresman, 1930.
- Russell, C., *Standard Tests. A Handbook for the Classroom Teacher*. Ginn and Company, 1930.
- Smith, H. L., and Wright, W. W., *Tests and Measurements*. Silver Burdett, 1928.
- Symonds, P. M., "Choice of items for a test on the basis of difficulty," *Journal of Educational Psychology*, XX (1929) 481-93.
- Symonds, P. M., "Factors influencing test reliability," *Journal of Educational Psychology*, IX (1928) 73-87.
- Thurstone, L. L., "Scale construction with weighted items," *Journal of Educational Psychology*, XIX (1928) 441-453.
- Thurstone, L. L., "A scoring method for mental tests," *Psychological Bulletin*, XVI (1919) 235-240.
- Tyler, Ralph W., "A generalized technique for conducting achievement tests," *Educational Research Bulletin*, X, No. 8, Ohio State University, 1931.
- Walker, Helen M., "Certain mathematical questions suggested by the true-false test," *American Mathematical Monthly*, XXXIV (1927) 503-507.
- Waples, D., "The best answer exercises as a teaching device," *Journal of Educational Research*, XV, 10-21.
- Weidemann, C. C., "Determinate and indeterminate information," *Journal of Educational Method*, VII (1928) 126-27.

BIBLIOGRAPHY OF TESTS

- Weidemann, C. C., *How to Construct the True-False Examination*. Teachers College, Contributions to Education, No. 225, Columbia University, 1926.
- Weidemann, C. C., "Limitations of the true-false statement," *Journal of Educational Method*, VII (1928) 214-215.
- Worcester, D. A., "Prevalent errors in new-type examinations," *Journal of Educational Research*, XVIII (1928) 18-48-52.

The exact titles of the reports of the Commission cannot yet be definitely announced. The following, therefore, must be regarded as tentative.

1. A Charter for the Social Sciences in the Schools, by Charles A. Beard, formerly Professor of Politics, Columbia University.
2. An Introduction to the History of the Social Sciences in Schools, by Henry Johnson, Professor of History, Teachers College, Columbia University.
3. Citizens' Organizations and the Civic Training of Youth, by Bessie Louise Pierce, Associate Professor of American History, University of Chicago.
4. Tests and Measurements in the Social Sciences, by Truman L. Kelley, Professor of Education, Harvard University, and A. C. Krey, Professor of History, University of Minnesota.
5. Geography in Relation to the Social Sciences, by Isaiah Bowman, Director, American Geographical Society of New York, with a special study by Rose Clark, Nebraska Wesleyan University.
6. Civic Education in the United States, by Charles E. Merriam, Professor of Political Science, University of Chicago.
7. The Nature of the Social Sciences, by Charles A. Beard, formerly Professor of Politics, Columbia University.
8. Educational Administration as Social Policy, by Jesse H. Newlon, Director, Lincoln School, Teachers College, Columbia University.
9. The Social Foundations of Education, by George S. Counts, Professor of Education, Teachers College, Columbia University, and Charles A. Beard.
10. The Social Ideas of American Educators, by Merle E. Curti, Professor of History at Smith College.
11. The Social Studies as School Subjects, by Rolla M. Tryon, Professor of the Teaching of History, University of Chicago.
12. Freedom of Teaching in the Schools, by Howard K. Beale, formerly Professor of History at Bowdoin College.
13. Methods of Instruction in the Social Sciences, by Ernest Horn, Professor of Education, University of Iowa.
14. A Social Process Approach to Curriculum-making in the Social Studies, by Leon C. Marshall, Institute for the Study of Law, Johns Hopkins University.

15. **The Selection and Training of the Teacher**, by William C. Bagley, Professor of Education, Teachers College, Columbia University; Guy Stanton Ford, Professor of History and Dean of the Graduate School, University of Minnesota; and others.
16. **Conclusions and Recommendations of the Commission.**

Other volumes or reports have been projected, dealing with the history of the social ideas of American educational leaders and with the problem of freedom of teaching or the growth of tolerance in the teaching of those subjects.

It is also planned to publish a number of miscellaneous studies, many of them of an exploratory character, which have been made for the committee in one connection or another.

To be published by Charles Scribner's Sons.